DAGSTUHL
REPORTS

Volume 5, Issue 11, November 2015

Dagstuhl Reports, Vol. 5, Issue 11 ISSN 2192-5283

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
http://www.dagstuhl.de/dagpub/2192-5283

Publication date
April, 2016

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at `http://dnb.d-nb.de`.

Digital Object Identifier: 10.4230/DagRep.5.11.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Editorial Office
Marc Herbstritt *(Managing Editor)*
Jutka Gasiorowski *(Editorial Assistance)*
Dagmar Glaser *(Editorial Assistance)*
Thomas Schillo *(Technical Assistance)*

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
http://www.dagstuhl.de/dagrep

Report of Dagstuhl Seminar 15451

Verification of Evolving Graph Structures

Edited by

Parosh Aziz Abdulla[1], Fabio Gadducci[2], Barbara König[3], and Viktor Vafeiadis[4]

1 **Uppsala University, SE,** parosh@it.uu.se
2 **University of Pisa, IT,** gadducci@di.unipi.it
3 **Universität Duisburg-Essen, DE,** barbara_koenig@uni-due.de
4 **MPI-SWS – Kaiserslautern, DE,** viktor@mpi-sws.org

Abstract

This report documents the programme and the outcome of Dagstuhl Seminar 15451 "Verification of Evolving Graph Structures".

The aim was to bring together researchers from different communities (shape analysis, separation logic, graph transformation, verification of infinite-state systems) who are interested in developing techniques for the analysis of graph manipulations, i.e., methods that are able to handle the challenges that arise in current verification problems.

Apart from scientific talks, the programme also included four tutorial talks and four working groups, which are summarized in this report.

Seminar November 1–6, 2015 – http://www.dagstuhl.de/15451
1998 ACM Subject Classification D.2.4 Software/Program Verification, F.3.1 Specifying and Verifying and Reasoning about Programs, F.4.2 Grammars and Other Rewriting Systems
Keywords and phrases dynamic systems, graph transformation, graphs, heap analysis, separation logic, shape analysis, static analysis, verification
Digital Object Identifier 10.4230/DagRep.5.11.1
Edited in cooperation with Christina Jansen and Eugenio Orlandelli

1 Summary

Parosh Aziz Abdulla
Fabio Gadducci
Barbara König
Viktor Vafeiadis

Despite significant progress in recent years, verification still remains a challenging task for hardware and software systems. A particularly complex verification problem is the analysis of graph-like structures that may modify their topology during runtime. The main reason for the difficulty is that some features give rise to infinite state spaces. Examples include variables ranging over unbounded domains, timing constraints, dynamic process creation, heap manipulation, multi-threading, and dynamically allocated data structures. An additional source of complication is that the underlying graphs may be continuously evolving. There is no a priori bound on the size of the graphs that may arise when modelling the run of a program, and the graph shapes may change during a given execution.

Verification of Evolving Graph Structures, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 1–28
Editors: Parosh Aziz Abdulla, Fabio Gadducci, Barbara König, and Viktor Vafeiadis
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This challenge has prompted several successful lines of research, developing novel techniques such as shape analysis, separation logic, forest automata, and several graph transformation-based approaches. Although specialized tools have been developed in each application area, a considerable amount of effort is needed to develop uniform frameworks that yield efficient yet general solutions.

This seminar brought together researchers interested in developing precise and scalable techniques for the analysis of graph manipulations, i.e., techniques that are able to handle the challenges that arise in current verification problems. These challenges require novel developments and the combination of techniques from a wide range of different areas including model checking and dynamic and static program analysis. By creating collaboration opportunities we hope to substantially increase the size of the systems that can be tackled and the precision of analysis that can be achieved.

Hence the main goal of this seminar was to enhance common understanding and cross-fertilization, highlighting connections among the approaches via tutorials and working groups, with the explicit purpose to enhance interaction. Discussion topics included:

- the definition of uniform frameworks in which to integrate methods for graph analysis that have been proposed by the different research communities;
- the development of new abstraction techniques for pushing the state-of-the-art of graph algorithms in program verification and model checking applications; and
- the identification of research areas in which the analysis of graph manipulation may play an important role, such as the analysis of security protocols, social networks, adaptive networks, and biological systems.

We invited four representatives of the different communities to give tutorial talks in order to introduce fundamental concepts and techniques. Specifically, the following four tutorial talks took place on the first day of the seminar:

- Tomas Vojnar: Shape Analysis via Symbolic Memory Graphs and Its Application for Conversion of Pointer Programs to Container Programs
- Giorgio Delzanno: Graphs in Infinite-State Model-Checking
- Arend Rensink: Verification Techniques for Graph Rewriting
- Viktor Vafeiadis: Separation Logic

On Tuesday and Thursday we organized the following working groups in order to discuss more specific topics which were of interest to a substantial part of the participants:

- Benchmarks and Application Domains
- Specification Languages for Graphs
- Ownership
- Graph Rewriting for Verification

The organizers would like to thank all the participants and speakers for their inspiring talks and many interesting discussions. Furthermore we would like to acknowledge Christina Jansen and Eugenio Orlandelli who helped to write and prepare this report. A special thanks goes to the Dagstuhl staff who were a great help in organizing this seminar.

2 Table of Contents

Summary

Overview of Talks

Summary of Working Groups

3 Overview of Talks

3.1 Verification of Dynamic Register Automata

Mohamed-Faouzi Atig (Uppsala University, SE)

Joint work of Parosh Aziz Abdulla; Mohamed Faouzi Atig; Ahmet Kara; Othmane Rezine
Main reference P. A. Abdulla, M. F. Atig, A. Kara, O. Rezine, "Verification of Dynamic Register Automata," in Proc. of the 34th Int'l Conf. on Foundation of Software Technology and Theoretical Computer Science (FSTTCS'14), LIPIcs, Vol. 29, pp. 653–665, Schloss Dagstuhl, 2014.
URL http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2014.653

We consider the verification problem for Dynamic Register Automata (DRA). DRA extend classical register automata by process creation. In this setting, each process is equipped with a finite set of registers in which the process IDs of other processes can be stored. A process can communicate with processes whose IDs are stored in its registers and can send them the content of its registers. The state reachability problem asks whether a DRA reaches a configuration where at least one process is in an error state. We will first show that this problem is in general undecidable. This result holds even when we restrict the analysis to configurations where the maximal length of the simple paths in their underlying (un)directed communication graphs are bounded by some constant. Then we will introduce the model of degenerative DRA which allows non-deterministic reset of the registers. We will prove that for every given DRA, its corresponding degenerative one has the same set of reachable states. While the state reachability of a degenerative DRA remains undecidable, we will show that the problem becomes decidable with nonprimitive recursive complexity when we restrict the analysis to strongly bounded configurations, i.e. configurations whose underlying undirected graphs have bounded simple paths. Finally, we will consider the class of strongly safe DRA, where all the reachable configurations are assumed to be strongly bounded. We show that for strongly safe DRA, the state reachability problem becomes decidable.

3.2 Verification Linearizability for SLL-based Concurrent Data Structures

Parosh Aziz Abdulla (Uppsala University, SE), Bengt Jonsson (Uppsala University, SE), and Cong-Quy Trinh

We present a framework for automated verification of linearizability for concurrent data structures that implement sets, stacks, and queues. We use a specification formalism for linearization policies which allows the user to specify complex patterns including non-fixed linearization points. We define abstraction techniques that allow to make the size of the data domain and the number of threads finite. In order to reason about dynamically allocated memory, we use a combination of shape graphs and thread-modular reasoning. Based on our method, we have verified linearizability for a number of algorithms., including all implementations of concurrent sets, stacks, and queues based on singly-linked lists that are known to us from the literature.

3.3 A Causal View on Non-Interference

Paolo Baldan (University of Padova, IT)

Joint work of Alessandro Beggiato; Alberto Carraro
Main reference P. Baldan, A. Carraro, "A Causal View on Non-Interference". Fundam. Inform. 140(1):1–38, 2015.
URL http://dx.doi.org/10.3233/FI-2015-1243

The concept of non-interference has been introduced to characterise the absence of undesired information flows in a computing system. Although it is often explained referring to an informal notion of causality – the activity involving the part of the system with higher level of confidentiality should not cause any observable effect at lower levels – it is almost invariably formalised in terms of interleaving semantics. In this talk, focusing on Petri nets, we discuss the possibility of providing a causal characterisations of non-interference based on a true concurrent semantics. The investigation can have a conceptual interest – as it clarifies the relation between causality and non-interference, and a practical value as the verification phase can take advantage of partial order techniques.

3.4 Observable Under-Approximations

Aiswarya Cyriac (Uppsala University, SE)

Joint work of Cyriac Aiswarya; Paul Gastin; K. Narayan Kumar
Main reference A. Cyriac, P. Gastin, K. N. Kumar, "Controllers for the Verification of Communicating Multi-Pushdown Systems," in Proc. of the 25th Int'l Conf. on Concurrency Theory (CONCUR'14), LNCS, Vol. 8704, pp. 297–311, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-662-44584-6_21

An under-approximation is observable/controllable/monitorable/diagnosable if it can be decided in run-time whether the current behaviour has exceeded the under-approximation. Behaviours of interest are graphs which, along a run, are monotonously increasing, by means of adding new vertices and edges. We illustrate the notion of observable under-approximation by an example on message sequence charts. We conclude the short presentation with some open questions. Is bounded tree-width observable as an under-approximation, even in special classes of graphs with an underlying linear order and uniformly bounded degree? Which classical under-approximations are observable?

3.5 Graphs in Infinite-State Model Checking (Tutorial)

Giorgio Delzanno (University of Genova, IT)

Joint work of P. A. Abdulla; M. F. Atig; N. Ben Henda; N. Bertrand; B. König; A. Rezine; O. Rezine; A. Sangnier; J. Stückrath; R. Traverso; G. Zavattaro
Main reference G. Delzanno, A. Sangnier, G. Zavattaro, "Parameterized Verification of Ad Hoc Networks," in Proc. of the 21th Int'l Conf. on Concurrency Theory (CONCUR'10), LNCS, Vol. 6269, pp. 313–327, Springer, 2010.
URL http://dx.doi.org/10.1007/978-3-642-15375-4_22

We present a survey on the application of graph theory in the field of infinite-state and parameterized verification. We consider different types of formalisms like transition systems, Petri nets, automata, and rewriting and discuss verification methods based on abstractions and symbolic state exploration.

References

1 P. A. Abdulla, M. F. Atig, and O. Rezine. Verification of directed acyclic ad hoc networks. In *FMOODS/FORTE*, pages 193–208, 2013.

2 P. A. Abdulla, G. Delzanno, and A. Rezine. Approximated parameterized verification of infinite-state processes with global conditions. *Formal Methods in System Design*, 34(2):126–156, 2009.

3 P. A. Abdulla, G. Delzanno, and A. Rezine. Automatic verification of directory-based consistency protocols with graph constraints. *Int. J. Found. Comput. Sci.*, 22(4), 2011.

4 P. A. Abdulla, G. Delzanno, O. Rezine, A. Sangnier, and R. Traverso. On the verification of timed ad hoc networks. In *FORMATS'11*, volume 6604 of *LNCS*, pages 256–270. Springer, 2011.

5 P. A. Abdulla, N. Ben Henda, G. Delzanno, and A. Rezine. Handling parameterized systems with non-atomic global conditions. In *VMCAI'08*, volume 4905 of *LNCS*, pages 22–36. Springer, 2008.

6 N. Bertrand, G. Delzanno, B. König, A. Sangnier, and J. Stückrath. On the decidability status of reachability and coverability in graph transformation systems. In *RTA*, pages 101–116, 2012.

7 N. Bertrand, P. Fournier, and A. Sangnier. Playing with probabilities in reconfigurable broadcast networks. In *FoSSaCS*, pages 134–148, 2014.

8 G. Delzanno, C. Di Giusto, M. Gabbrielli, C. Laneve, and G. Zavattaro. The *kappa*-lattice: Decidability boundaries for qualitative analysis in biological languages. In *CMSB*, pages 158–172, 2009.

9 G. Delzanno, A. Sangnier, and R. Traverso. Parameterized verification of broadcast networks of register automata. In *RP*, pages 109–121, 2013.

10 G. Delzanno, A. Sangnier, R. Traverso, and G. Zavattaro. On the complexity of parameterized reachability in reconfigurable broadcast networks. In *FSTTCS'12*, volume 18 of *LIPIcs*, pages 289–300. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2012.

11 G. Delzanno, A. Sangnier, and G. Zavattaro. Parameterized verification of ad hoc networks. In *CONCUR'10*, volume 6269 of *LNCS*, pages 313–327. Springer, 2010.

12 G. Delzanno, A. Sangnier, and G. Zavattaro. On the power of cliques in the parameterized verification of ad hoc networks. In *FOSSACS'11*, volume 6604 of *LNCS*, pages 441–455. Springer, 2011.

13 G. Delzanno, A. Sangnier, and G. Zavattaro. Verification of ad hoc networks with node and communication failures. In *FORTE/FMOODS'12*, volume 7273 of *LNCS*, pages 235–250. Springer, 2012.

14 G. Delzanno and R. Traverso. Decidability and complexity results for verification of asynchronous broadcast networks. In *LATA*, pages 238–249, 2013.

15 G. Ding. Subgraphs and well quasi ordering. *J. of Graph Theory*, 16(5):489–502, 1992.

16 E. A. Emerson and K. S. Namjoshi. On model checking for non-deterministic infinite-state systems. In *LICS'98*, pages 70–80. IEEE Computer Society, 1998.

17 E. Allen Emerson and V. Kahlon. Parameterized model checking of ring-based message passing systems. In *Computer Science Logic, 18th International Workshop, CSL 2004, 13th Annual Conference of the EACSL, Karpacz, Poland, September 20-24, 2004, Proceedings*, pages 325–339, 2004.

18 S. M. German and A. P. Sistla. Reasoning about systems with many processes. *J. ACM*, 39(3):675–735, 1992.

19 S. Joshi and B. König. Applying the graph minor theorem to the verification of graph transformation systems. In *CAV'08*, volume 5123 of *LNCS*, pages 214–226. Springer, 2008.

20 K. S. Namjoshi and R. J. Trefler. Uncovering symmetries in irregular process networks. In *VMCAI*, pages 496–514, 2013.

21 M. Saksena, O. Wibling, and B. Jonsson. Graph grammar modeling and verification of ad hoc routing protocols. In *TACAS*, pages 18–32, 2008.

3.6 PSYNC: A Partially Synchronous Language for Fault-Tolerant Distributed Algorithms

Cezara Dragoi (IST Austria – Klosterneuburg, AT), Damien Zufferey, and Tom Henzinger

Fault-tolerant distributed algorithms play an important role in many critical/high-availability applications. These algorithms are notoriously difficult to implement correctly, due to asynchronous communication and the occurrence of faults, such as the network dropping messages or computers crashing. We introduce PSYNC, a domain specific language based on the Heard-Of model, which views asynchronous faulty systems as synchronous ones with an adversarial environment that simulates asynchrony and faults by dropping messages. We define a runtime system for PSYNC that efficiently executes on asynchronous networks. We formalize the relation between the runtime system and PSYNC in terms of observational refinement. This high-level synchronous abstraction introduced by PSYNC simplifies the design and implementation of fault-tolerant distributed algorithms and enables automated formal verification. We have implemented an embedding of PSYNC in the SCALA programming language with a runtime system for partially synchronous networks. We show the applicability of PSYNC by implementing several important fault-tolerant distributed algorithms and we compare the implementation of consensus algorithms in PSYNC against implementations in other languages in terms of code size, runtime efficiency, and verification.

3.7 Symbolic Abstract Data Types

Constantin Enea (University of Paris VII, FR)

Formal specification is a vital ingredient to scalable verification of software systems. In the case of efficient implementations of concurrent objects like atomic registers, queues, and locks, symbolic formal representations of their abstract data types (ADTs) enable efficient modular reasoning, decoupling clients from implementations. Writing adequate formal specifications, however, is a complex task requiring rare expertise. In practice, programmers write reference implementations as informal specifications.

In this work we demonstrate that effective symbolic ADT representations can be automatically generated from the executions of reference implementations. Our approach exploits two key features of naturally-occurring ADTs: violations can be decomposed into a small set of representative patterns, and these patterns manifest in executions with few operations. By identifying certain algebraic properties of naturally-occurring ADTs, and exhaustively sampling executions up to a small number of operations, we generate concise symbolic ADT representations which are complete in practice, enabling the application of efficient symbolic verification algorithms without the burden of manual specification. Furthermore, the concise ADT violation patterns we generate are human-readable, and can serve as useful, formal documentation.

3.8 From Decision Procedures to Full Model-Checking: the MCMT Experience

Silvio Ghilardi (University of Milan, IT)

In this talk, we briefly report both experience and case studies in the development of our logic-based models checker, called MCMT, 'Model Checker Modulo Theories' (see http://users.mat.unimi.it/users/ghilardi/mcmt). During past years, many people (besides the author of the present contribution) contributed to implementation, theoretical advances or experiments; among them, let us mention F. Alberti, R. Bruttomesso, A. Carioni, E. Nicolini, A. Orsini, E. Pagani, S. Ranise, N. Sharygina, D. Zucchelli.

The basic idea in the development of MCMT is the revisitation, in a declarative perspective, of classic results concerning well structured transition systems (WSTS) [2]. The WSTS framework is a formal framework covering a large class of systems and their evolution; in concrete applications, WSTS arise from finitely presented models of rather simple logical theories. These theories are array theories obtained from the combination of a theory for process 'topology' and of theories for (local and shared) data. The notion of an *array-based system* formalizes this intuition [18, 19]. In array-based systems backward search can be implemented in a purely symbolic way and a model-checker exploiting this idea can rely on state-of-the-art SMT-solvers to discharge the proof-obligations needed for fixpoint and safety tests (no ad hoc data structures need to be invented). Monotonic abstraction techniques [3, 4, 5, 1] can be implemented declaratively too by using syntactic constructions like quantifier instantiations and quantifiers relativizations [11].

The framework of array-based system is quite flexible and can be adapted to cope with various kinds of distributed [19], timed [17, 16] and fault-tolerant systems [9, 10]. Both abstraction [6, 8] and acceleration [14] techniques can be integrated with it, making the approach quite suitable for dealing also with array-manipulating sequential programs [7, 12]. For future, we expect even more progress taking advantage from recent advances in the decision procedures concerning quantified fragments of array theories [15, 20, 13, 14].

References

1 P. A. Abdulla. Forcing monotonicity in parameterized verification: From multisets to words. In *Proceedings of SOFSEM '10*, pages 1–15. Springer-Verlag, 2010.

2 P. A. Abdulla, K. Cerans, B. Jonsson, and Y.-K. Tsay. General decidability theorems for infinite-state systems. In *Proc. of LICS*, pages 313–321, 1996.

3 P. A. Abdulla, G. Delzanno, N. B. Henda, and A. Rezine. Regular model checking without transducers. In *TACAS*, volume 4424 of *LNCS*, pages 721–736, 2007.

4 P. A. Abdulla, G. Delzanno, and A. Rezine. Parameterized verification of infinite-state processes with global conditions. In *CAV*, pages 145–157, 2007.

5 P. A. Abdulla, N. B. Henda, G. Delzanno, and A. Rezine. Handling parameterized systems with non-atomic global conditions. In *Proc. of VMCAI*, volume 4905 of *LNCS*, pages 22–36, 2008.

6 F. Alberti, R. Bruttomesso, S. Ghilardi, S. Ranise, and N. Sharygina. Lazy Abstraction with Interpolants for Arrays. In *LPAR-18*, pages 46–61, 2012.

7 F. Alberti, R. Bruttomesso, S. Ghilardi, S. Ranise, and N. Sharygina. SAFARI: SMT-Based Abstraction for Arrays with Interpolants. In *CAV*, pages 679–685, 2012.

8 F. Alberti, R. Bruttomesso, S. Ghilardi, S. Ranise, and N. Sharygina. An extension of lazy abstraction with interpolation for programs with arrays. *Formal Methods in System Design*, pages 63–109, 2014.

9 F. Alberti, S. Ghilardi, E. Pagani, S. Ranise, and G. P. Rossi. Automated support for the design and validation of fault tolerant parameterized systems – a case study. In *Proc. of AVOCS*, 2010.

10 F. Alberti, S. Ghilardi, E. Pagani, S. Ranise, and G. P. Rossi. Brief announcement: Automated support for the design and validation of fault tolerant parameterized systems – a case study. In *DISC*, pages 392–394, 2010.

11 F. Alberti, S. Ghilardi, E. Pagani, S. Ranise, and G.P. Rossi. Universal guards, relativization of quantifiers, and failure models in model checking modulo theories. *JSAT*, 8(1/2):29–61, 2012.

12 F. Alberti, S. Ghilardi, and N. Sharygina. Booster : an acceleration-based verification framework for array programs. In *ATVA*, pages 18–23, 2014.

13 F. Alberti, S. Ghilardi, and N. Sharygina. Decision procedures for flat array properties. In *TACAS*, pages 15–30, 2014.

14 F. Alberti, S. Ghilardi, and N. Sharygina. A new acceleration-based combination framework for array properties. In *FroCoS*, 2015.

15 A.R. Bradley, Z. Manna, and H.B. Sipma. What's decidable about arrays? In *VMCAI*, pages 427–442, 2006.

16 R. Bruttomesso, A. Carioni, S. Ghilardi, and S. Ranise. Automated Analysis of Parametric Timing Based Mutual Exclusion Protocols. In *NASA Formal Methods Symposium*, 2012.

17 A. Carioni, S. Ghilardi, and S. Ranise. MCMT in the land of parameterized timed automata. In *In proc. of VERIFY*, 2010.

18 S. Ghilardi, E. Nicolini, S. Ranise, and D. Zucchelli. Towards SMT Model-Checking of Array-based Systems. In *Proc. of IJCAR*, LNCS, 2008.

19 S. Ghilardi and S. Ranise. Backward Reachability of Array-based Systems by SMT solving: Termination and Invariant Synthesis. *LMCS*, 6(4), 2010.

20 P. Habermehl, R. Iosif, and T. Vojnar. A logic of singly indexed arrays. In *LPAR*, pages 558–573, 2008.

3.9 Invariant Checking for Graph Transformation: Applications & Open Challenges

Holger Giese (Hasso-Plattner-Institut – Potsdam, DE) and Leen Lambers (Hasso-Plattner-Institut – Potsdam, DE)

Graph transformation can be used as a formal foundation for modeling different kinds of evolving graph structures. In particular, we present two application domains, cyber-physical systems (CPS) and model-driven engineering (MDE), where graph transformation has been used successfully to formally model different scenarios. Furthermore, we employ inductive invariant checking for graph transformation [1] as a verification technique for the different scenarios of the two domains. In the CPS domain the invariance of important safety properties can be shown. In the MDE domain, behavior preservation of model transformations can be reduced to invariant checking [2]. We give an overview of how invariant checking has been applied in these two domains on different scenarios. We present the strengths and weaknesses of this verification technique and conclude with some open challenges.

References

1 Basil Becker, Dirk Beyer, Holger Giese, Florian Klein, and Daniela Schilling. Symbolic invariant verification for systems with dynamic structural adaptation. In *Proc. of the* 28^th *International Conference on Software Engineering (ICSE), Shanghai, China.* ACM Press, 2006.

2 Holger Giese and Leen Lambers. Towards automatic verification of behavior preservation for model transformation via invariant checking. In *Proceedings of International Conference on Graph Transformation (ICGT'12)*, volume 7562 of *LNCS*, pages 249–263. Springer, 2012.

3.10 Approaching the Coverability Problem Continuously

Christoph Haase (ENS – Cachan, FR)

Joint work of Michael Blondin; Alain Finkel; Serge Haddad
Main reference M. Blondin, A. Finkel, C. Haase, S. Haddad, "Approaching the Coverability Problem Continuously," arXiv:1510.05724v2 [cs.LO], 2016.
URL http://arxiv.org/abs/1510.05724v2

The coverability problem for Petri nets plays a central role in the verification of concurrent shared-memory programs. However, its high EXPSPACE-complete complexity poses a challenge when encountered in real-world instances. In this talk, I will present a new approach to this problem which is primarily based on applying forward coverability in continuous Petri nets as a pruning criterion inside a backward-coverability framework. A cornerstone of the approach is the efficient encoding of a recently developed polynomial-time algorithm for reachability in continuous Petri nets into SMT. The effectiveness of the approach is demonstrated on standard benchmarks from the literature, which shows that it decides significantly more instances than any existing tool and is in addition often much faster, in particular on large instances.

3.11 Dijkstra-style Verification of Graph Programs

Annegret Habel (Universität Oldenburg, DE)

Joint work of Annegret Habel; Karl-Heinz Pennemann; Hendrik Radke; Nils Erik Flick

We investigate Dijkstra-style verification of graph programs relative to several notions of graph conditions (nested, recursively nested, M, HR*) and show: (1) For all notions of graph conditions, there is a transformation Wp such that for every graph program P and every postcondition post, Wp(P,post) is a weakest precondition of P relative to post. (2) For nested and recursively nested graph conditions, there is a semi-decider for the implication problem. In particular, the theorem prover for nested graph conditions is much better than the theorem provers for first-order graph formulas getting the transformed graph condition as input.

References

1 Annegret Habel and Karl-Heinz Pennemann. Correctness of high-level transformation systems relative to nested conditions.Mathematical Structures in Computer Science, 19:245–296, 2009.

2 Annegret Habel and Hendrik Radke. Expressiveness of graph conditions with variables. Electronic Communications of the EASST, 30, 2010.

3 Nils Erik Flick. On correctness of graph programs relative to recursively nested conditions. In Graph Computation Models (GCM 2015), volume 1403, pages 97–112. CEUR-WS.org, 2015.

3.12 Modelling Evolving Graph Structures by Differential Equations

Reiko Heckel (University of Leicester, GB)

Joint work of Reiko Heckel; Mudhafar Hussein
Main reference Mudhafar Hussein, Reiko Heckel, Vincent Danos, Pawel Sobocinski, "Modelling Adaptive Networks: The Case of the Petrified Voters," Electronic Communications of the EASST, Vol. 67, pp. 1–12, 2014.
URL http://dx.doi.org/10.14279/tuj.eceasst.67.950

From a stochastic graph transformation system modelling an evolving network it is possible to derive a system of differential equations describing the average evolution of the network. The key concept is an approximation of complex patterns such as they appear in rules' left- and right-hand sides, including negative application conditions and attribute constraints, by combinations of simpler ones. Such approximations is correct in the sense that, for large random graphs where occurrences of patterns and attribute values are independent, they converge towards the actual number of occurrences of the complex patterns. We illustrate this approach by an example and discuss how these assumptions can be validated using simulations.

References

1 Mudhafar Hussein, Reiko Heckel, Vincent Danos, Pawel Sobocinski. *Modelling Adaptive Networks: The Case of the Petrified Voters.* Proceedings of the 13th International Workshop on Graph Transformation and Visual Modeling Techniques (GTVMT 2014)

3.13 A Graph-Based Semantics Workbench for Concurrent Asynchronous Programs

Alexander Heußner (Universität Bamberg, DE) and Chris Poskitt (ETH Zürich, CH)

Joint work of Claudio Corrodi; Alexander Heußner; Christopher M. Poskitt

This talk presents a pathway to a "semantics workbench", with which multiple alternative and possibly contradicting semantics of state-of-the-art concurrency abstractions can be formalised, analysed, and compared. We also raise some (new) fundamental research questions in the areas of graph transformation systems and verification.

3.14 Inverse Monoid of Higher Dimensional Strings

David Janin (University of Bordeaux, FR)

Halfway between graph transformation theory and inverse semigroup theory, we define higher dimensional strings as bi-deterministic graphs with distinguished sets of input roots and output roots. We show that these generalized strings can be equipped with an associative product so that the resulting algebraic structure is an inverse semigroup. Its natural order is shown to capture existence of root preserving graph morphism. A simple set of generators is characterized. As a subsemigroup example, we show how all finite grids are finitely generated. Finally, simple additional restrictions on products lead to the definition of subclasses with decidable Monadic Second Order (MSO) language theory.

References

1 D. Janin. Inverse monoids of higher-dimensional strings. In *Int. Col. on Theor. Aspects of Comp. (ICTAC)*, volume 9399 of *LNCS*, 2015.

3.15 Verifying Pointer Programs using Graph Grammars

Christina Jansen (RWTH Aachen, DE), Joost-Pieter Katoen (RWTH Aachen, DE), Christoph Matheja, and Thomas Noll (RWTH Aachen, DE)

Joint work of Jonathan Heinen; Christina Jansen; Joost-Pieter Katoen; Christoph Matheja; Thomas Noll
Main reference J. Heinen, C. Jansen, J.-P. Katoen, T. Noll, "Juggrnaut: Using graph grammars for abstracting unbounded heap structures," Formal Methods in System Design, 47(2):159–203, 2015.
URL http://dx.doi.org/10.1007/s10703-015-0236-1

This talk presents an abstraction framework for heap data structures. It employs graph grammars, more precisely context-free hyperedge replacement grammars. Our approach aims at extending finite-state verification techniques to handle pointer-manipulating programs operating on complex dynamic data structures that are potentially unbounded in their size. We will see which subset of hyperedge replacement grammars provides sound abstractions and briefly elaborate on its relation to Separation Logic. In addition, a small tool comparison comprising our prototypical tool Juggrnaut as well as each a tool from the area of shape analysis, Separation Logic and general graph transformation is presented.

3.16 Bounded Time-Stamping for Message-passing Systems: Beyond Channel Bounds

Narayan Kumar Krishnan (Chennai Mathematical Institute, IN)

Joint work of Cyriac Aiswarya; Paul Gastin; K. Narayan Kumar

Consider distributed systems consisting of a number of processes communicating with each other by sending messages via FIFO channels. It is crucial for such systems that every process can maintain deterministically the latest information about other processes. To do so, any process p, upon receiving a message from a process q, should determine for every process r, whether the latest event on r that p knows of is more recent than the latest event on r that q knows of. Solving this problem, while storing and exchanging only a bounded amount of information, is very challenging and not always possible. This is known as the gossip problem. A solution to this is the key to solving numerous important problems on distributed systems. We provide a solution that simplifies an existing algorithm for this problem and also extends it to a richer class, going beyond a priori channel bounds.

3.17 Spatio-Temporal Model Checking

Michele Loreti (University of Firenze, IT)

Joint work of Vincenzo Ciancia; Stephen Gilmore; Gianluca Grilletti; Diego Latella; Michele Loreti; Mieke Massink

The interplay between process behaviour and spatial aspects of computation has become more and more relevant in Computer Science, especially in the field of collective adaptive systems, but also, more generally, when dealing with systems distributed in physical space.

Traditional verification techniques are well suited to analyse the temporal evolution of programs; properties of space are typically not explicitly taken into account. We propose a methodology to verify properties depending upon physical space. We define an appropriate logic, stemming from the tradition of topological interpretations of modal logics, dating back to earlier logicians such as Tarski, where modalities describe neighbourhood and surrounding. We lift the topological definitions to a more general setting, also encompassing discrete, graph-based structures. A spatial extension of the global model checking algorithm of the temporal logic CTL is also presented. More precisely, we add to CTL the new spatial operators. The interplay of space and time permits one to define complex spatio-temporal properties.

References

1 Specifying and Verifying Properties of Space V. Ciancia, D. Latella, M. Loreti, M. Massink IFIP TCS 2014, Lecture Notes in Computer Science, 8705, pp. 222–235, 2014.

2 Spatio-Temporal Model-Checking Of Vehicular Movement In Public Transport Systems V. Ciancia, S. Gilmore, G. Grilletti, D. Latella, M. Loreti and M. Massink. Submitted for journal publication, 2014.

3 A spatio-temporal model-checker V. Ciancia, G. Grilletti, D. Latella, M. Loreti and M. Massink. VERY* 2015, Lecture Notes in Computer Science, to appear.

4 Qualitative and Quantitative Monitoring of Spatio-Temporal Properties L. Nenzi, L. Bortolussi, V. Ciancia, M. Loreti and M. Massink. RV 2015, Lecture Notes in Computer Science, 9333, pp. 21–37, 2015.

3.18 Pointer Race Freedom

Roland Meyer (TU Kaiserslautern, DE)

Joint work of Frédéric Haziza; Lukas Holik; Roland Meyer; Sebastian Wolff
URL http://arxiv.org/abs/1511.00184

We propose a novel notion of pointer race for concurrent programs manipulating a shared heap. A pointer race is an access to a memory address which was freed, and it is out of the accessor's control whether or not the cell has been re-allocated. We establish two results. (1) Under the assumption of pointer race freedom, it is sound to verify a program running under explicit memory management as if it was running with garbage collection. (2) Even the requirement of pointer race freedom itself can be verified under the garbage-collected semantics. We then prove analogues of the theorems for a stronger notion of pointer race needed to cope with performance-critical code purposely using racy comparisons and even racy dereferences of pointers. As a practical contribution, we apply our results to optimize a thread-modular analysis under explicit memory management. Our experiments confirm a speed-up of up to two orders of magnitude.

3.19 Modular Analysis of Concurrent Pointer Programs Using Graph Grammars

Thomas Noll (RWTH Aachen, DE), Christina Jansen (RWTH Aachen, DE), and Jens Katelaan

Programs with shared-memory concurrency are inherently difficult to get right: they are prone to all the memory-related errors that are familiar from the single-threaded setting, such as null pointer dereferences and unintended aliasing. In addition, the possible interference between parallel execution threads gives rise to new classes of errors, such as data races. As thread interleaving is nondeterministic in nature and heap-manipulating programs generally have an unbounded state space due to dynamic memory allocation, the application of formal methods is challenging in this setting.

In this talk we develop a static analysis for proving properties such as shape invariants, absence of null pointer dereferences, as well as data-race freedom of programs with fork-join parallelism. To this end, we develop a formal semantics based on hypergraphs and access permissions, and derive an abstract interpretation that uses hyperedge replacement grammars to safely approximate the program's semantics. The result is a fully automatic, thread-modular analysis for proving the above properties in the presence of recursive data structures and dynamic (possibly recursive) thread creation.

3.20 On Graphical Logics for Reasoning about Graph Properties

Fernando Orejas (UPC – Barcelona, ES)

Joint work of Leen Lambers; Marisa Navarro; Fernando Orejas; Elvira Pino

By graphical logics, we mean logics whose formulas are not text, but they consist of graphs and graph morphisms, that are used to express graph properties. In this presentation, I will review previous work on this area and, moreover, I will present the main problems found when extending the logic to allow for the specification of the existence of paths in given graphs. In particular, I will present a proof calculus for this extension that is shown to be sound and it is conjectured to be complete.

References

1 F. Orejas, H. Ehrig and U. Prange: *Reasoning with Graph Constraints*, Form. Asp. Computing 2010.
2 A. Rensink: *Representing First-Order Logic Using Graphs.* ICGT 2004: 319–335
3 A. Habel, K. H. Pennemann: *Correctness of high-level transformation systems relative to nested conditions.* Math. Structures in Comp. Science 2009.
4 K. H. Pennemann: *Development of Correct Graph Transformation Systems*, Ph. D. Thesis, 2009.
5 L. Lambers, F. Orejas: *Tableau-Based Reasoning for Graph Properties.* ICGT 2014.
6 M. Navarro, F. Orejas, E. Pino: *Satisfiability of Constraint Specifications on XML Documents.* Festschrift José Meseguer (2015).

3.21 Interactive Verification of Parameterized Systems

Oded Padon (Tel Aviv University, IL)

Joint work of Ken McMillan; Oded Padon; Mooly Sagiv

Verification of infinite-state and parameterized systems is a long standing research goal. Automated verification tools for such systems often solve an intractable to undecidable problem, so some failures of automation are unavoidable. This work is an an attempt to combine user interaction with automated verification heuristics. We use the decidable EPR fragment of FOL to obtain predictability of the automated analysis, and engage the user to help the system generalize from counter-examples to induction. The user interaction is obtained via graphical visualization, and interactive heuristics. By combining powerful invariant inference heuristics with user interaction, we hope to make verification of infinite-state and parameterized systems more practical.

3.22 Hoare-Style Verification for GP 2

Detlef Plump (University of York, GB) and Christopher M. Poskitt (ETH Zürich, CH)

Joint work of Christopher M. Poskitt; Detlef Plump
Main reference C. M. Poskitt, D. Plump, "Hoare-Style Verification of Graph Programs", Fundamenta Informaticae, 118(1–2): 35–175, 2012.
URL http://dx.doi.org/10.3233/FI-2012-708

GP 2 is an experimental non-deterministic programming language for solving problems on graphs and graph-like structures. The language is based on graph transformation rules, allowing visual programming at a high level of abstraction. We introduce GP 2 and present a Hoare-style proof system for assertional reasoning about programs. The pre- and postconditions of our calculus are nested graph conditions with extensions for properties of attributes and monadic second-order graph structure. This allows us to reason about global properties of graphs, such as 2-colourability, existence of paths, or connectedness. Our proof system is sound with respect to the operational semantics of GP 2.

References

1 C. M. Poskitt, D. Plump, Verifying monadic second-order properties of graph programs, in: Proc. International Conference on Graph Transformation (ICGT 2014), Vol. 8571 of LNCS, Springer, 2014, pp. 33–48.

2 C. M. Poskitt, D. Plump, Hoare-style verification of graph programs, Fundamenta Informaticae 118 (1-2) (2012) 135–175.

3.23 Verification Techniques for Graph Rewriting (Tutorial)

Arend Rensink (University of Twente, NL)

This tutorial paints a high-level picture of the concepts involved in verification of graph transformation systems. We distinguish three fundamentally different application scenarios for graph rewriting: (1) as grammars (in which case we are interested in the language, or set, of terminal graphs for a fixed start graph); (2) as production systems (in which case we are interested in the relation between start and terminal graphs); or (3) as behavioural specifications (in which case we are interested in the transition system as a whole). We then list some types of questions one might want to answer through verification: confluence and termination, reachability, temporal properties, or contractual properties. Finally, we list some techniques that can help in providing answers: model checking, unfolding, assertional reasoning, and abstraction.

3.24 Refining Orderings for Parameterized Verification

Ahmed Rezine (Linköping University, SE)

Joint work of Ahmed Rezine; Zeinab Ganjei; Yu-Fang Chen; Parosh Abdulla; Giorgio Delzanno; Petru Elese; Zebo Peng
Main reference Z. Ganjei, A. Rezine, P. Eles, Z. Peng, "Abstracting and counting synchronizing processes," in Proc. of the 16th Int'l Conf. on Verification, Model Checking, and Abstract Interpretation (VMCAI'15), LNCS, Vol. 8931, pp. 227–244, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-662-46081-8_13
Main reference Z. Ganjei, A. Rezine, P. Eles, Z. Peng, "Lazy Constrained Monotonic Abstraction," in Proc. of the 17th Int'l Conf. on Verification, Model Checking, and Abstract Interpretation (VMCAI'16), LNCS, Vol. 9583, pp. 147–165, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-662-49122-5_7

Multi-threaded programs may synchronise in subtle ways. For instance, they can use integer variables to count the number of threads satisfying some property in order to implement dynamic barriers or to organise their interleaved execution. We address the problem of automatically establishing deadlock freedom and safety in general for multi-threaded programs generating an arbitrary number of concurrent processes. For this purpose, we explain how we leverage on simple techniques to derive "counting invariants", i.e., invariants that relate the number or processes in a given location to the values of the program variables. We use these invariants and leverage on predicate abstraction techniques in order to generate non-monotonic counter machine reachability problems that faithfully capture the correctness of the safety property.

We describe how we check reachability for non-monotonic counter machines. The idea is to localise the refinement of well quasi orderings in order to allow for a decidable reachability analysis on possibly infinite abstractions that are well structured wrt. these orderings. The orderings can be refined based on obtained false positives in a CEGAR like fashion. This allows for the verification of systems that are not monotonic and are hence inherently beyond the reach of classical well-structured-systems-based analysis techniques. Unlike classical lazy predicate abstraction, we show the feasibility of the approach even for systems with infinite control. Our heuristics are applicable both in backward and in forward as shown by our experiments.

References

1 Z. Ganjei, A. Rezine, P. Eles, and Z. Peng. Lazy Constrained Monotonic Abstraction. VMCAI 2016.
2 Z. Ganjei, A. Rezine, P. Eles, and Z. Peng. Abstracting and counting synchronizing processes. VMCAI 2015.
3 Alastair F. Donaldson, Alexander Kaiser, Daniel Kroening, Thomas Wahl. Symmetry-Aware Predicate Abstraction for Shared-Variable Concurrent Programs. CAV 2011.
4 Parosh Aziz Abdulla, Yu-Fang Chen, Giorgio Delzanno, Frédéric Haziza, Chih-Duo Hong, Ahmed Rezine. Constrained Monotonic Abstraction: A CEGAR for Parameterized Verification. CONCUR 2010.
5 Parosh Aziz Abdulla, Giorgio Delzanno, Ahmed Rezine. Parameterized Verification of Infinite-State Processes with Global Conditions. CAV 2007.

3.25 Shape Analysis for Unstructured Sharing

Xavier Rival (ENS – Paris, FR)

Joint work of Huisong Li; Bor-Yuh Evan Chang; Xavier Rival

Shape analysis aims to infer precise structural properties of imperative memory states and has been applied heavily to verify safety properties on imperative code over pointer-based data structures. It is often applied to dynamic structures such as lists and trees, that can be summarised using inductive predicates. Unfortunately, data structures with unstructured sharing, such as graphs, are challenging to describe and reason about in such frameworks. In particular, when the sharing is unstructured, it cannot be described inductively and in a purely local manner. In this talk, we will describe a global abstraction of sharing based on set-valued variables that when integrated with inductive definitions enables the specification and shape analysis of structures with unstructured sharing.

3.26 Local Strategies in Selective Broadcast Networks

Arnaud Sangnier (University of Paris VII, FR)

Joint work of Nathalie Bertrand; Paulin Fournier
Main reference N. Bertrand, P. Fournier, A. Sangnier, "Distributed Local Strategies in Broadcast Networks," in Proc. of the 26th Int'l Conf. on Concurrency Theory (CONCUR'15), LIPIcs, Vol. 42, pp. 44-57, Schloss Dagstuhl, 2015.
URL http://dx.doi.org/10.4230/LIPIcs.CONCUR.2015.44

We study the problems of reaching a specific control state, or converging to a set of target states, in networks with a parameterized number of identical processes communicating via broadcast. To reflect the distributed aspect of such networks, we restrict our attention to executions in which all the processes must follow the same local strategy that, given their past performed actions and received messages, provides the next action to be performed.

3.27 Compositional Reasoning and Symmetry For Dynamic Protocol Analysis

Richard Trefler (University of Waterloo, CA)

Joint work of Zarrin Langari; Kedar Namjoshi; Richard Trefler

We consider the problem of analyzing programs of several processes that operate over an underlying network. Model checking and other program analysis engines applied to such programs may suffer from the state explosion problem, in which the number of global states to be analyzed is exponential in the number of component processes that compose the system. Compositional reasoning techniques decompose the problem of reasoning about the global system states to a local problem that reasons about the component processes one by one. Symmetry reduction techniques allow many similar processes to be considered all at once by choosing a single representative process as an instance of any one of the symmetric individual processes. We show how to tailor the notion of local process symmetry to apply in the context of compositional analysis. This allows local symmetry reduction techniques to be applied in cases where notions of global symmetry are not applicable. Utilizing abstractions on the local processes we can then apply local symmetry and compositional analysis techniques to locally symmetric protocols; parametrized families of locally symmetric protocols; and dynamic protocols, where the number of processes and their underlying connection network is not fixed during program execution.

References

1 Kedar S. Namjoshi, Richard J. Trefler: Loop Freedom in AODVv2. FORTE 2015: 98–112
2 Kedar S. Namjoshi, Richard J. Trefler: Analysis of Dynamic Process Networks. TACAS 2015: 164–178
3 Kedar S. Namjoshi, Richard J. Trefler: Uncovering Symmetries in Irregular Process Networks. VMCAI 2013: 496–514
4 Kedar S. Namjoshi, Richard J. Trefler: Local Symmetry and Compositional Verification. VMCAI 2012: 348–362
5 Zarrin Langari, Richard J. Trefler: Symmetry for the Analysis of Dynamic Systems. NASA Formal Methods 2011: 252–266
6 Zarrin Langari, Richard J. Trefler: Formal Modeling of Communication Protocols by Graph Transformation. FM 2006: 348–363

3.28 Separation Logic (Tutorial)

Viktor Vafeiadis

Separation logic is an extension of Hoare logic introduced by Reynolds, O'Hearn and others [1], that is suitable for reasoning about programs manipulating heap-allocated data structures. The tutorial gave an overview of separation logic. It covered both sequential and concurrent separation logic [2, 11], and presented an opinionated view of the pros and cons of separation logic.

References

1 Reynolds, J., Separation logic: A logic for shared mutable data structures. In *LICS 2002* (2002), pp. 55–74

2 Brookes, S., *A semantics for concurrent separation logic*, Theor. Comput. Sci. **375** (2007), pp. 227–270.

3 O'Hearn, P. W., *Resources, concurrency and local reasoning*, Theor. Comput. Sci. **375** (2007), pp. 271–307.

3.29 Shape Analysis via Symbolic Memory Graphs and Its Application for Conversion of Pointer Programs to Container Programs (Tutorial)

Tomas Vojnar (Brno University of Technology, CZ)

Joint work of Kamil Dudka; Lukas Holik; Petr Peringer; Marek Trtik; Tomas Vojnar
Main reference K. Dudka, P. Peringer, T. Vojnar, "Byte-Precise Verification of Low-Level List Manipulation," in Proc. of the 20th Int'l Symposium on Static Analysis (SAS'13), LNCS, Vol. 7935, pp. 215–237, Springer, 2013.
URL http://dx.doi.org/10.1007/978-3-642-38856-9_13
Main reference K. Dudka, L. Holik, P. Peringer, M. Trtik, T. Vojnar, "From Low-Level Pointers to High-Level Containers," in Proc. of 17th Int'l Conf. on Verification, Model Checking, and Abstract Interpretation (VMCAI'16), LNCS, Vol. 9583, pp. 431–452, Springer, 2016.
URL http://dx.doi.org/10.1007/978-3-662-49122-5_21

In the talk, we briefly overview the main principles of shape analysis based on symbolic memory graphs (SMGs) as implemented in the Predator analyser for C programs with low-level pointer operations (such as pointer arithmetic, address alignment, or block operations). Subsequently, we present a novel application of shape analysis for converting pointer programs to programs using high-level (list) containers.

3.30 Automating Separation Logic Using SMT

Thomas Wies (New York University, US)

Joint work of Ruzica Piskac; Thomas Wies; Damien Zufferey

Separation logic (SL) has gained widespread popularity as a formal foundation of tools that analyze and verify heap-manipulating programs. Its great asset lies in its assertion language, which can succinctly express how data structures are laid out in memory, and its discipline of local reasoning, which mimics human intuition about how to prove heap programs correct.

While the succinctness of separation logic makes it attractive for developers of program analysis tools, it also poses a challenge to automation: separation logic is a nonclassical logic that requires specialized theorem provers for discharging the generated proof obligations. SL-based tools therefore implement their own tailor-made theorem provers for this task. However, these theorem provers are not robust under extensions, e.g., involving reasoning about the data stored in heap structures.

I will present an approach that enables complete combinations of decidable separation logic fragments with other theories in an elegant way. The approach works by reducing

SL assertions to first-order logic. The target of this reduction is a decidable fragment of first-order logic that fits well into the SMT framework. That is, reasoning in separation logic is handled entirely by an SMT solver. We have implemented our approach in the GRASShopper tool and used it successfully to verify interesting data structures.

3.31 Shape and Content

Florian Zuleger (TU Wien, AT)

Joint work of Diego Calvanese; Tomer Kotek; Mantas Simkus; Helmut Veith; Florian Zuleger

Pointers in programs serve two different purposes: (1) Storage of information; pointers are used to build data structures. (2) Semantic information; pointers relate different data items to each other. While the program analysis community has spent considerable effort on analyzing the shape of pointer structures, much less effort has been spent on the analysis of data structure content and the relationship between data items. In this talk I will argue that two-variable logic with counting (C2) is an interesting choice for content anlaysis as it can describe UML-like properties, model pointers and express weakest preconditions of pointer programs [1]. I will discuss extensions of C2 that can express data structures such as lists and trees [2]. Further I will present a combination of C2 with MSO over graphs with bounded tree-width; the resulting logic allows to describe complex data structures and is still decidable [1].

References

1 Tomer Kotek, Helmut Veith, Florian Zuleger: Monadic second order finite satisfiability and unbounded tree-width. arXiv preprint arXiv:1505.06622 (2015)

2 Tomer Kotek, Mantas Simkus, Helmut Veith, Florian Zuleger: Extending ALCQIO with Trees. LICS 2015: 511-522

3 Diego Calvanese, Tomer Kotek, Mantas Simkus, Helmut Veith, Florian Zuleger: Shape and Content – A Database-Theoretic Perspective on the Analysis of Data Structures. IFM 2014: 3-17

4 Summary of Working Groups

4.1 Working Group: Benchmarks and Application Domains

The starting point of the workgroup was a quick introduction to exemplary benchmarks in the different domains of the workgroup participants, e.g. a car platooning case study and a particular routing protocol for mobile ad-hoc networks (AODV routing).

Soon the group agreed that a benchmark collection is of great importance for progress (on tools and approaches) in the area of graph-based analysis and verification. The library SMT-LIB substantiates the above claim and was brought up as an example. Here the problem statement is provided by this widely accepted library and it is common that tool developers in the SMT community provide parsers to automatically transfer the SMT-LIB input problems into a form of input their own tools accept. However, the field of graph-based analysis and verification is a lot more diverse.

Consequently, it was discussed if fixing a language in which problems can be stated in our setting can be realised in a reasonable manner at all. On the one hand restricting to an input language would increase the efficiency for reusing benchmarks drastically, while manual translation is laborious and error-prone. On the other hand in this diverse setting there will always be features that a fixed language does not cover.

Stating the problem of a benchmark and the checked properties, e.g. by providing reference models, was accepted as a solution/compromise.

As outcome of the discussion the benchmarks workgroup decided to take action in setting up an (initial) spreadsheet to collect benchmarks. During the session it was filled by benchmarks that participants suggested to serve as exemplary entries.

Benchmarks Spreadsheet

The following points are identified to be the goal of the benchmarks collection:

- state problems, not solutions
- obstacles to access and edit should be very low
- for now a more or less unsorted collection is targeted, categorisation is provided by tags users can enlist

Concept Discussion

After first setting up an initial spreadsheet for the collection, details on its contents are discussed in the second session.

The group agreed on collecting information on the benchmark's name and a (very) short description. Moreover, a category with tags like 'distributed' or 'seq. heap' and the underlying graph dynamics (e.g. adding/removing edges/nodes) as well as difficulties of the benchmark and interesting properties are requested. Optionally, it is possible to provide the originator and a reference model. For additional information, such as example models for the benchmark or solutions/papers that tackle it, a column 'References' is added. Here one can provide a link to a subfolder where the information (e.g. in the form other further spreadsheets, plain text, ...) is found.

To guide new users, explaining the purpose of the benchmarks collection and the policy regarding the quality of the entries, an additional Readme-document is made available.

The second session was completed by a short demo on a reference model provided as a graph transformation system (in a file format that is accepted by the GROOVE tool).

Access

The benchmarks spreadsheet is made available to the public via a google spreadsheet accessible at

http://tiny.cc/egbd.

The spreadsheet was presented to all participants of the seminar with the appeal to spread the word.

4.2 Working Group: Specification Languages for Graphs

The working group was formed in order to compare, discuss and assess various specification formalism for graphs. Specifically, it discussed the following questions: (i) Which logics and/or specification languages are used by the different communities to specify (possibly infinite) sets of graphs? (ii) How do they compare with respect to their expressiveness? (iii) What are possible applications? (iv) What are desirable properties of such specification languages?

In order to answer questions (i) and (ii) the working group compiled the following list of specification languages and discussed their expressiveness:

- Recursively enumerable graph languages
- Context-free graph languages, generated by hyperedge-replacement grammars
- Monadic second-order logic, with strong ties to recognizable graph languages
- First-order logic, equivalent to nested graph conditions
- Separation logic
- Forest automata
- Type graphs
- DATALOG
- Propositional Dynamic Logic
- Spatial Logic for Closure Spaces

Afterwards the group talked about item (iii) and discussed applications in verification such as reachability analysis, bounded model-checking and counterexample-guided abstraction refinement.

Finally the working group turned its attention to point (iv) and came up with the following list of important and/or desirable properties of such specification languages (with respect to decidability and complexity):

- Membership test
- Satisfiability
- Entailment
- Computation of post-/preconditions (for certain transformations)
- Invariant checking/closure under rewriting
- Computation of interpolants and interpolant-like notions
- Expressiveness (which properties can be expressed?)
- Ease of use (i.e., a specification language should specify a set of graphs that is expressible in the underlying formalism in a way that is both modular and easy to read/manipulate)
- Closure properties (wrt. negation, disjunction, conjunction, concatenation, . . .)
- Framed inference ($F\star? \models G\star?$)
- Widening/approximation

In the end the main open question that arose is to find the best balance between expressiveness and complexity (for problems such as satisfiability, entailment and computation of post-/preconditions). Naturally, this has to be tailored according to the the application. The central open problem that has emerged during the discussion is the need of a better classification/overview of available specification languages with respect to their relative expressiveness and their properties.

4.3 Working Group: Ownership

The working group discussed the various techniques that are used to reason about how the ownership of various resources is distributed to different components of a software system.

Ownership disciplines became very important for programming languages, such as C and C++, where memory is explicitly managed. In such setting, it is very important that unused memory cells are deallocated exactly once. Deallocating a given memory cell multiple times may corrupt the contents of memory, while not deallocating a memory cell constitutes a space leak. The ownership of a memory cell is a very useful concept, because it defines the component responsible for deallocating the cell. Moreover, since ownership of a memory cell is required for a function to access the cell, ownership systems ensure memory safety; that is, absence of memory errors, such as accessing a deallocated memory cell.

Ownership is also particularly useful in the concurrent setting, where each thread requires ownership of a memory cell in order to access it. By doing so, one can ensure that a program does not have any data races. This means that one does not need to worry about any weak memory effects of the hardware, because weak memory models typically ensure that race-free programs have interleaving semantics.

An important ownership technique that the group discussed at length was *separation logic* [13]. In separation logics, the unit of ownership is a memory cell, but small pieces of ownership can be combined together and abstracted away using *abstract predicates* [12]. Extensions of separation logic with fractional and counting *permissions* [2, 1] enable also partial ownership of a memory cell, which is sufficient for reading, but not writing to it.

Concurrent separation logic (CSL) [11] is an extension of separation logic that handles interleaving concurrency. CSL allows a component to increase its ownership capabilities by acquiring a lock, and decrease them by releasing a lock. Carefully chosen invariants for locks enable even *ownership transfer* from one thread to another via a lock synchronization. This idea has also been used in the weak memory setting using *relaxed separation logic* [16]. Further, by careful ownership transfer of fractional permissions, one can encode various other ownership disciplines, such as a single writer with multiple atomic readers. Ownership is a very important component of almost all modern concurrent program logics such as RGSep [17], LRG [6], CAP [5], CaReSL [15], TADA [4], IRIS [8].

Besides separation logic, there are other ways of ensuring an ownership discipline. There is a rich literature on *ownership type systems*, which are very useful for describing well-structured forms of ownership. A nice survey about ownership types is given by Clarke et al. [3]. Unfortunately, there is no corresponding survey for separation logic techniques. Another more recent idea is that of *implicit dynamic frames* [14]. Moreover, there are several tools based on these ideas. Examples are Dafny [9], VeriFast [7], and Viper [10].

References

1 R. Bornat, C. Calcagno, P. W. O'Hearn, and M. J. Parkinson. Permission accounting in separation logic. In *POPL*, pages 259–270. ACM, 2005.

2 J. Boyland. Checking interference with fractional permissions. In *10th SAS*, volume 2694 of *LNCS*, pages 55–72. Springer, 2003.

3 D. Clarke, J. Östlund, I. Sergey, and T. Wrigstad. Ownership types: A survey. In *Aliasing in Object-Oriented Programming*, volume 7850 of *LNCS*, pages 15–58. Springer, 2013.

4 P. da Rocha Pinto, T. Dinsdale-Young, and P. Gardner. Tada: A logic for time and data abstraction. In *ECOOP*, volume 8586 of *LNCS*, pages 207–231. Springer, 2014.

5 T. Dinsdale-Young, M. Dodds, M. P. Philippa Gardner, and V. Vafeiadis. Concurrent abstract predicates. In *ECOOP'10*, Lecture Notes in Computer Science. Springer, 2010.

6 X. Feng. Local rely-guarantee reasoning. In *POPL*, pages 315–327, 2009.

7 B. Jacobs, J. Smans, P. Philippaerts, F. Vogels, W. Penninckx, and F. Piessens. VeriFast: A powerful, sound, predictable, fast verifier for C and Java. In *NASA Formal Methods*, volume 6617 of *LNCS*, pages 41–55. Springer, 2011.

8 R. Jung, D. Swasey, F. Sieczkowski, K. Svendsen, A. Turon, L. Birkedal, and D. Dreyer. Iris: Monoids and invariants as an orthogonal basis for concurrent reasoning. In *POPL*, pages 637–650. ACM, 2015.

9 K. R. M. Leino. Dafny: An automatic program verifier for functional correctness. In *LPAR (Dakar)*, volume 6355 of *LNCS*, pages 348–370. Springer, 2010.

10 P. Müller, M. Schwerhoff, and A. J. Summers. Viper: A verification infrastructure for permission-based reasoning. In *VMCAI*, volume 9583 of *Lecture Notes in Computer Science*, pages 41–62. Springer, 2016.

11 P. W. O'Hearn. Resources, concurrency and local reasoning. *Theoretical Computer Science*, 375(1-3):271–307, 2007.

12 M. J. Parkinson and G. M. Bierman. Separation logic and abstraction. In *POPL*, pages 247–258. ACM, 2005.

13 J. C. Reynolds. Separation logic: A logic for shared mutable data structures. In *LICS*, pages 55–74. IEEE Computer Society, 2002.

14 J. Smans, B. Jacobs, and F. Piessens. Implicit dynamic frames. *ACM Trans. Program. Lang. Syst.*, 34(1):2, 2012.

15 A. Turon, D. Dreyer, and L. Birkedal. Unifying refinement and Hoare-style reasoning in a logic for higher-order concurrency. In *ICFP*. ACM, 2013.

16 V. Vafeiadis and C. Narayan. Relaxed separation logic: A program logic for C11 concurrency. In *OOPSLA 2013*, pages 867–884. ACM, 2013.

17 V. Vafeiadis and M. Parkinson. A marriage of rely/guarantee and separation logic. In L. Caires and V. T. Vasconcelos, editors, *CONCUR*, volume 4703 of *LNCS*, pages 256–271. Springer, 2007.

4.4 Working Group: Graph Rewriting for Verification

Graph rewriting is a rule-based formalism for the transformation of graphs. Rules are local and replace a left-hand side with a right-hand side, taking also some embedding information into account. In his invited talk the day before Arend Rensink distinguished between three types of graph transformation systems: (i) grammars (in which case we are interested in the language, or set, of terminal graphs for a fixed start graph); (ii) production systems (in which case we are interested in the relation between start and terminal graphs); and (iii) behavioural specifications (in which case we are interested in the transition system as a whole).

The aim of the working group was to discuss applications of graph rewriting in verification. In principle, two possible applications came to mind: (1) using graph rewriting as an auxiliary technique to support other verification methods; (2) developing methods in order to specifically analyze and verify graph transformation systems.

As a first case study (presented by Christoph Haase) the group discussed a production system to support entailment checks in separation logic. In separation logic it is not possible to do entailment checks directly, but only after having reduced the formulas to normal form. The group studied a reduction rule that collapses two nodes in case there are certain distinct paths.

Such rules have a non-local flavour and can not be straightforwardly expressed in classical graph rewriting. The group discussed possible extensions that would be desirable in order to be able to express such rules.

The second case study (presented by Barbara König) was a behavioural specification (a termination detection protocol), where the aim is to verify whether an invalid configuration (termination has been declared, but there are still active processes) can be reached. For this specific task it is for instance possible to use a backward analysis technique based on well-structured transition systems. For finite systems also classical model-checking can be used, but many interesting systems do not satisfy finiteness.

A third case study (presented by Richard Trefler) involved the verification of telephone communication. In this case we have a graph that is evolving and de-evolving and models the communication structure, and the protocol can be described by local rule-based transformations. Verification is concerned with problems such as loop-checking, involving compositional reasoning (in the sense that the description of the global behaviour is given by means of the description of the local ones).

The working group concluded by discussing the question of which concepts in graph rewriting could be helpful in the design of verification techniques. One answer was the concept of morphisms that allow us to precisely describe the relation between two graphs or the occurrence of a graph within a host graph.

Figure 1 The seminar group during the excursion on Wednesday afternoon.

Participants

- Mohamed-Faouzi Atig
Uppsala University, SE
- Parosh Aziz Abdulla
Uppsala University, SE
- Peter Backes
Universität des Saarlandes, DE
- Paolo Baldan
University of Padova, IT
- Ahmed Bouajjani
University of Paris VII, FR
- Andrea Corradini
University of Pisa, IT
- Aiswarya Cyriac
Uppsala University, SE
- Giorgio Delzanno
University of Genova, IT
- Cezara Dragoi
IST Austria –
Klosterneuburg, AT
- Constantin Enea
University of Paris VII, FR
- Javier Esparza
TU München, DE
- Fabio Gadducci
University of Pisa, IT
- Silvio Ghilardi
University of Milan, IT
- Holger Giese
Hasso-Plattner-Institut –
Potsdam, DE
- Christoph Haase
ENS – Cachan, FR
- Annegret Habel
Universität Oldenburg, DE
- Reiko Heckel
University of Leicester, GB
- Alexander Heußner
Universität Bamberg, DE
- Lukas Holik
Brno Univ. of Technology, CZ
- David Janin
University of Bordeaux, FR
- Christina Jansen
RWTH Aachen University, DE
- Bengt Jonsson
Uppsala University, SE
- Joost-Pieter Katoen
RWTH Aachen, DE
- Barbara König
Universität Duisburg-Essen, DE
- Tomer Kotek
TU Wien, AT
- Narayan Kumar Krishnan
Chennai Mathematical Inst., IN
- Leen Lambers
Hasso-Plattner-Institut –
Potsdam, DE
- Michele Loreti
University of Firenze, IT
- Roland Meyer
TU Kaiserslautern, DE
- Thomas Noll
RWTH Aachen, DE
- Fernando Orejas
UPC – Barcelona, ES
- Eugenio Orlandelli
CIS, IT
- Oded Padon
Tel Aviv University, IL
- Detlef Plump
University of York, GB
- Chris Poskitt
ETH Zürich, CH
- Arend Rensink
University of Twente, NL
- Ahmed Rezine
Linköping University, SE
- Leila Ribeiro
Federal University of Rio Grande
do Sul, BR
- Xavier Rival
ENS – Paris, FR
- Arnaud Sangnier
University of Paris VII, FR
- Richard Trefler
University of Waterloo, CA
- Viktor Vafeiadis
MPI-SWS – Kaiserslautern, DE
- Tomas Vojnar
Brno Univ. of Technology, CZ
- Thomas Wies
New York University, US
- Florian Zuleger
TU Wien, AT

Report of Dagstuhl Perspectives Workshop 15452

Artifact Evaluation for Publications

Edited by

Bruce R. Childers[1], Grigori Fursin[2], Shriram Krishnamurthi[3], and Andreas Zeller[4]

1 **University of Pittsburgh, US,** childers@cs.pitt.edu
2 **cTuning - Cachan, FR,** grigori.fursin@ctuning.org
3 **Brown University - Providence, US,** sk@cs.brown.edu
4 **Universität des Saarlandes, DE,** zeller@cs.uni-saarland.de

Abstract

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 15452 “Artifact Evaluation for Publications”. This Perspectives Workshop conveyed several stakeholders in artifact evaluation from different communities to assess how artifact evaluation is working and make recommendations to the computer systems research community about several issues with the process.

Seminar November 1–4, 2015 – http://www.dagstuhl.de/15452
1998 ACM Subject Classification H.3.7 Digital Libraries, H.3.5 Online Information Services, D.2.9 Management
Keywords and phrases Computer systems, artifacts, reproducibility, archive
Digital Object Identifier 10.4230/DagRep.5.11.29

1 Executive Summary

Bruce R. Childers
Grigori Fursin
Shriram Krishnamurthi
Andreas Zeller

Computer systems researchers have developed numerous artifacts that encompass a broad collection of software tools, benchmarks, and data sets. These artifacts are used to prototype innovations, evaluate trade-offs and analyze implications. Unfortunately, methods used in the evaluation of computing system innovation are often at odds with sound science and engineering practice. The ever-increasing pressure to publish more and more results poses an impediment to *accountability*, which is a key component of the scientific and engineering process. Experimental results are not usually disseminated with sufficient metadata (i.e., software extensions, data sets, benchmarks, test cases, scripts, parameters, etc.) to achieve repeatability and/or reproducibility. Without this information, issues surrounding trust, fairness and building on and comparing with previous ideas becomes problematic. Efforts in various computer systems research sub-communities, including programming languages/compilers, computer architecture, and high-performance computing, are underway to address the challenge.

This Dagstuhl Perspectives Workshop (PW) brought together stakeholders of associated CSR sub-communities to determine synergies and to identify the most promising directions

Artifact Evaluation for Publications, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 29–35
Editors: Bruce R. Childers, Grigori Fursin, Shriram Krishnamurthi, and Andreas Zeller
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and mechanisms to push the broader community toward accountability. The PW assessed current efforts, shared what does and doesn't work, identified additional processes, and determined possible incentives and mechanisms. The outcomes from the workshop, including recommendations to catalyze the community, are separately documented in an associated Dagstuhl Manifesto.

2 Table of Contents

3 Goals

Before the workshop, the organizers identified several goals to engage and drive the event. These goals were:

1. Assess the state of current efforts to achieve accountability, including successes and why these worked, and the impediments being and likely to be faced;
2. Identify strategies and incentives to engage the community and raise expectation for higher experimental quality and accountability;
3. Identify the shape of the most promising approaches for the technical challenges posed by building open-access repositories and associated services;
4. Identify ways to leverage, combine and coordinate existing and new efforts in PL/compilers and software engineering, high-performance computing and computer architecture;
5. Develop recommendations to lead the community to better artifacts and accountable experimental results.

4 Topics

To address the goals, the PW was be organized around four topics with several questions for participants to consider for each topic. Rather than hold formal talks, the workshop was arranged as a round-table discussion of the topics. The agenda was intentionally informal to allow plenty of time for significant discussion. Participant ideas, hunches, concerns, thoughts, and challenges were every bit as welcome as any particular concrete thing the participants had done. During the discussion, the intent was not to decisively answer the questions – that can only be done with full community involvement over time – but rather to give direction of where answers may be found and how to get there. All of the participants contributed to the discussion. Our informal format was quite successful in creating a lively and productive environment to work through the issues on artifact evaluation.

The specific outcomes (recommendations) from the workshop are described in the Dagstuhl Manifesto for the event. Below, we list the topics and the questions discussed.

4.1 Assessment of Existing Efforts

There are several initial efforts for accountability in CSR sub-communities for PL/compilers and software engineering, computer systems architecture, and HPC. These efforts were discussed to understand how well they are working and what issues are being faced. Some specific questions addressed were:

- What approaches taken by existing efforts have achieved the most traction, and what approaches have faced the most resistance?
- What obstacles have been and will likely be raised?
- What are the similarities and differences among sub-communities?
- What capabilities are currently missing?

4.2 Pushing the Community Further

To reap the benefits of accountability, researchers have to be moved to adopt expectations and procedures for this purpose. Community needs should first be determined, and then processes and infrastructures can be developed. Incentive is important – individuals must be motivated to participate. Several questions were discussed on how to catalyze the community to do more:

- What is the community's desired level of trust and leverage from accountability?
- How can the risk of over emphasis on building and evaluating artifacts and experiments be mitigated to avoid imposing too much hassle?
- How can the community be incentivised?
- How do we treat industrial artifacts?
- What is the interaction between evaluation of artifacts and paper acceptance?
- Should journals also participate in these processes?

4.3 Building and Sustaining a Community

Widespread community "buy-in" – from funding agencies, to program committees, to journal editors, to individual researchers – is necessary to establish and sustain accountability and associated processes and mechanisms. Accountability must become an inherent expectation for it to be effective and sustainable. Several questions were considered on how a community can be created to support artifact evaluation:

- How can the community be convinced that accountability is beneficial?
- What advocacy strategies will encourage adoption of processes and mechanisms for accountability?
- How can the community be facilitated toward overcoming concerns with privacy/accessibility?

4.4 Processes, Mechanisms and Repositories

Accountability relies on access to experimental details, which implies artifacts and associated metadata should be available. To leverage and compare with past innovation in the most effective way requires access to the original prototype implementation. Typically mundane issues with packaging and distribution become vital ones. This topic presented several questions for discussion about what technical capabilities are required:

- Do the AE processes which have been described (and which are working in their areas) work more generally across CS?
  - If so, what needs to be done to have them taken up by the whole community?
  - If not, how do we understand what needs to be changed?
- Is there more that we could do in terms of AE that would improve either trust or usefulness of artifacts?
- How should a repository and associated services be structured? A number of demos and examples were given of existing systems, but an structured analysis of differences and similarities is really required.
- What is a good taxonomy for the repositories and the services?
- How can artifacts, metadata and results be packaged as digital objects for a repository?
- Should journals also participate in these processes? For example, SCP/Elsevier is trying to take a path for artifacts, but the main issue here is "who owns what?" in scientific publications in general, and in artifacts in particular?

5 Demos

Several participants gave demos and/or discussions of systems that they have been developing or used for artifact evaluation. These demos included:

- Collective Knowledge, http://cTuning.org
- DataMill, https://uwaterloo.ca/embedded-software-group/datamill
- Open Curation for Computer Architecture, http://www.occamportal.org
- Parallel Workloads Archive, http://www.cs.huji.ac.il/labs/parallel/workload/
- Portable Database Files, http://www.vldb.org/pvldb/vol8/p1972-dittrich.pdf
- Multi2Sim Heterogeneous System Simulator, https://www.multi2sim.org/
- SPADE, https://github.com/ashish-gehani/spade
- TIRA – Evaluation as a Service, http://www.tira.io

Acknowledgments

We thank Carlo Ghezzi, Jan Vitek, Matthias Hauswirth, Camil Demetrescu, and the others that have worked with us on artifact evaluation for their input on the process.

Participants

- Bruce R. Childers
University of Pittsburgh –
Pittsburgh, USA
- Neil Chue Hong
Software Sustainability Institute –
Edinburgh, GB
- Tom Crick
Cardiff Metropolitan University –
Cardiff, GB
- Jack W. Davidson
University of Virginia –
Charlottesville, USA
- Camil Demetrescu
Sapienza University of Rome –
Rome, IT
- Roberto Di Cosmo
Univ. Paris-Diderot – Paris, FR
- Jens Dittrich
Universität des Saarlandes –
Saarbrücken, DE
- Dror Feitelson
The Hebrew University of
Jerusalem – Jerusalem, IL
- Sebastian Fischmeister
University of Waterloo –
Waterloo, CA
- Grigori Fursin
cTuning – Cachan, FR
- Ashish Gehani
SRI – Menlo Park, US
- Matthias Hauswirth
Univ. of Lugano – Lugano, CH
- Marc Herbstritt
Schloss Dagstuhl – Wadern, DE
- David R. Kaeli
Northeastern University –
Boston, US
- Shriram Krishnamurthi
Brown Univ. – Providence, US
- Anton Lokhmotov
Dividiti Ltd. – Cambridge, GB
- Martin Potthast
Bauhaus-Universität Weimar –
Weimar, DE
- Lutz Prechelt
FU Berlin, – Berlin, DE
- Petr Tuma
Charles University – Prague, CZ
- Michael Wagner
Schloss Dagstuhl – Wadern, DE
- Andreas Zeller
Universität des Saarlandes –
Saarbrücken, DE

Report of Dagstuhl Seminar 15461

Vision for Autonomous Vehicles and Probes

Edited by

Andrés Bruhn[1], Atsushi Imiya[2], Aleš Leonardis[3], and Tomas Pajdla[4]

1 **Universität Stuttgart, DE,** bruhn@vis.uni-stuttgart.de
2 **Chiba University, JP,** imiya@faculty.chiba-u.jp
3 **University of Birmingham, GB,** a.leonardis@cs.bham.ac.uk
4 **Czech Technical University Prague, CZ,** pajdla@cmp.fell.cvut.cz

Abstract

The vision-based autonomous driving and navigation of vehicles has a long history. In 2013, Daimler succeeded autonomous driving on a public drive way. Today, the Curiosity mars rover is sending video views from Mars to Earth. Computer vision plays a key role in advanced driver assistance systems (ADAS) as well as in exploratory and service robotics. Continuing topics of interest in computer vision are scene and environmental understanding using single- and multiple-camera systems, which are fundamental techniques for autonomous driving, navigation in unknown environments and remote visual exploration. Therefore, we strictly focuses on mathematical, geometrical and computational aspects of autonomous vehicles and autonomous vehicular technology which make use of computer vision and pattern recognition as the central component for autonomous driving and navigation and remote exploration.

Seminar November 8–13, 2015 – http://www.dagstuhl.de/15461

1998 ACM Subject Classification I.2.9 Robotics, I.2.10 Vision and Scene Understanding, I.4.5 Reconstruction, I.4.6 Segmentation, I.4.8 Scene Analysis

Keywords and phrases Vision-based autonomous driving and navigation, Exploratory rovers, Dynamic 3D scene understanding, Simultaneous localization and mapping, On-board algorithms

Digital Object Identifier 10.4230/DagRep.5.11.36

Edited in cooperation with Akihiro Torii (Tokyo Institute of Technology, JP) and Hayato Itoh (Chiba University, JP)

1 Executive Summary

Andrés Bruhn
Atsushi Imiya

Computer vision plays a key role in advanced driver assistance systems (ADAS) as well as in exploratory and service robotics. Visual odometry, trajectory planning for Mars exploratory rovers and the recognition of scientific targets in images are examples of successful applications. In addition, new computer vision theory focuses on supporting autonomous driving and navigation as applications to unmanned aerial vehicles (UAVs) and underwater robots. From the viewpoint of geometrical methods for autonomous driving, navigation and exploration, the on-board calibration of multiple cameras, simultaneous localisation and mapping (SLAM) in non-human-made environments and the processing of non-classical features are some of current problems. Furthermore, the adaptation of algorithms to long image sequences, image pairs with large displacements and image sequences with changing illumination is desired for robust navigation and exploration. Moreover, the extraction of non-verbal and graphical information from environments to remote driver assistance is required.

Based on these wide range of theoretical interests from computer vision for new possibility of practical applications of computer vision and robotics, 38 participants (excluding organisers) attended from variety of countries: 4 from Australia, 3 from Austria, 3 from Canada, 1 from Denmark, 11 from Germany, 1 from Greece, 1 from France, 3 from Japan, 4 from Spain, 2 from Sweden, 4 from Switzerland and 3 from the US.

The seminar was workshop style. The talks are 40 mins and 30 mins for young researchers and for presenters in special sessions. The talks have been separated into sessions on aerial vehicle vision, under water and space vision, map building, three-dimensional scene and motion understanding as well as a dedicated session on robotics. In these tasks, various types of autonomous systems such as autonomous aerial vehicles, under water robots, field and space probes for remote exploration and autonomous driving cars were presented. Moreover, applications of state-of-the-art computer vision techniques such as global optimization methods, deep learning approaches as well as geometrical methods for scene reconstruction and understanding were discussed. Finally, with Seminar 15462 a joint session on autonomous driving with leading experts in the field was organised.

The working groups are focused on "Sensing," "Interpretation and Map building" and "Deep leaning." Sensing requires fundamental methodologies in computer vision. Low-level sensing is a traditional problem in computer vision. For applications of computer-vision algorithms to autonomous vehicles and probes, reformulation of problems for various conditions are required. Map building is a growing area including applications to autonomous robotics and urban computer vision. Today, application to autonomous map generation involves classical SLAM and large-scale reconstruction from indoor to urban sizes. Furthermore, for SLAM on-board and on–line computation is required. Deep learning, which goes back its origin to '70s, is a fundamental tool for image pattern recognition and classification. Although the method showed significant progress in image pattern recognition and discrimination, for applications to spatial recognition and three-dimensional scene understanding, we need detailed discussion and developments.

Through talks-and-discussion and working-group discussion, the seminar clarified that for designing of platforms for visual interpretation and understanding of three-dimensional world around the system, machine vision provides fundamental and essential methodologies. There is the other methodology which uses computer vision as a sensing system for the acquisition of geometrical data and analysis of motion around cars. For these visual servo systems, computer vision is a part of the platform for intelligent visual servo system. The former methodology is a promising one to provide a fundamental platform which is common to both autonomous vehicles, which are desired for consumer intelligence, and probes, which are used for remote exploration.

2 Table of Contents

3 Overview of Talks

35 talks have been categorised as follows.

Vision for Mapping, Reconstruction and SLAM

Hayko Riemenschneider	Efficient Multi-view Semantic Segmentation
Michal Havlena	Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition
Antonios Gasteratos	Semantic Maps for High Level Robot Navigation
Daniel Cremers	Dense and Direct Methods for 3D Reconstruction and Visual SLAM
Vladyslav Usenko	Direct SLAM Techniques for Vehicle Localization and Autonomous Navigation
Akihiko Torii	Large-scale visual place recognition and online 3D reconstruction
Yasutaka Furukawa	Structured Indoor Modeling and/or Uncanny Valley for 3D Reconstruction
Torsten Sattler	The Limits of Pose Estimation in Very Large Maps

Vision for Aerial, Space and Underwater Robotics

Friedrich Fraundorfer	Drone Vision – Computer vision algorithms for drones
Davide Scaramuzza	From Frames to Events: Vision for High-speed Robotics
Takashi Kubota	Image based Navigation for Exploration Probe
Lazaros Nalpantidis	Stereo Vision for Future Autonomous Space Exploration Robots
Ben Huber	Planetary Robotic Vision Processing for NASA and ESA Rover Missions
Rafael Garcia	Underwater Vision: Robots that "see" beneath the surface

Vision for Scene Understanding

Jürgen Sturm	Tracking and Mapping in Project Tango
Raquel Urtasun	3D Scene Understanding for Autonomous Driving
Andreas Geiger	High-level Knowledge in Low-level Vision
Bernt Schiele	Towards 3D Scene Understanding

Vision for Motion Analysis

Cédric Demonceaux	Pose Estimation and 3D Segmentation using 3D Knowledge in Dynamic Environments
Florian Becker	Recursive Joint Estimation of Dense Scene Structure and Camera Motion in an Automotive Scenario
Johannes Berger	Second-Order Recursive Filtering on the Rigid-Motion Group $SE(3)$ Based on Nonlinear Observations from Monocular Videos
Michael Felsberg	Learning to Drive
Mikael Persson	Structure and Motion -Challenges and solutions for real time geometric estimation from video
Reihard Koch	Model-based Object and Deformation Tracking with Robust Global Optimization

Vision for Autonomous Driving and Robotics

Heiko Hirschmueller	Visual-Inertial Navigation for Mobile Robots
Sven Behnke	Semantic RGB-D Perception for Cognitive Robots
Darius Burschka	Robust Coupling of Perception to Actuation in Dynamic Environments
Juan Andrade-Cetto	Perception for Mobile Robotics
Niko Sünderhauf	Deep Learning for Visual Place Recognition and Online 3D Reconstruction
David Vázquez Bermudez	Learning See in a Virtual World
José Alvaerz	Real-world Semantic Segmentation
Steven Beauchemin	Vehicular Instrumentation for the Study of Driver Intent and Related Applications
Danil Prokhorov	Toward Highly Intelligent Automobiles
Andres Wendel	Realizing Self-Driving Car
Thomas Pock	Efficient Block Optimization Methods for Computer Vision

4 Talks Abstracts

4.1 Real-world Semantic Segmentation

José M. Alvarez (NICTA – Canberra, AU)

Semantic segmentation is a key low-level task to fully understand the environment of vehicle. Ideal semantic segmentation algorithms have four desirable properties: fast, robust, accurate and compact. Semantic segmentation methods must be fast to enable real-time high-level reasoning; Robust to operate at any-time in any weather conditions; Accurate enough to be reliable and, compact to facilitate functionalities in embedded platforms where power and resources are relevant. In this talk we present our recent work towards fast, robust and accurate semantic segmentation in embedded platforms.

4.2 Perception for Mobile Robotics

Juan Andrade-Cetto (UPC – Barcelona, ES)

Joint work of Andrade-Cetto, Juna; Andreasson, Henrik; Corominas, Andreu; Fleuret, Francois; Ila, Viorela; Lippiello, Vincenzo; Moreno-Noguer, Françesc; Peñate-Sanchez, Adriàn; Porta, Josep M.; Sanfeliu, Alberto; Teniente, Ernesto H.; Saarinen, Jari; Santamaria-Navarro, Ángel; Valencia, Rafael; Solá, Joan; Vallvè, Joan; Villamizar, Michael

In this talk I addressed several challenges on perception for mobile robotics. First I overview a pair of object detection and pose recognition algorithms that have the property of being very fast to compute. The first exploits the bootstrapping of very simple features on a boosting classifier. For the second one we propose the use of 3D annotated features. This allows camera pose estimation from low quality monocular images of a previously learned scene. Further in the talk I described our research on visual servoing for UAV manipulation and UAV odometry estimation using tight sensor data fusion. I then proceeded talking about mapping for mobile robots, and in particular about using principled information theoretic metrics to keep the map size tractable. These very same information metrics can also be used for optimal navigation, exploration and optimal sensor placement as explained also in the talk. I concluded the talk with the application of these research results for the autonomous driving of trucks and heavy load AGVs in cargo container terminals.

References

1 V. Ila, J.M. Porta and J. Andrade-Cetto, Information-based compact Pose SLAM, IEEE Transactions on Robotics, 26(1), 78–93, 2010.

2 V. Ila, J.M. Porta and J. Andrade-Cetto, Amortized constant time state estimation in Pose SLAM and hierarchical SLAM using a mixed Kalman-information filter, Robotics and Autonomous Systems, 59(5), 310–318, 2011.

3 A. Peñate-Sanchez, J. Andrade-Cetto and F. Moreno-Noguer, Exhaustive linearization for robust camera pose and focal length estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(10), 2387–2400, 2013.

4 A. Peñate-Sanchez, F. Moreno-Noguer, J. Andrade-Cetto and F. Fleuret, LETHA: Learning from high quality inputs for 3D pose estimation in low quality images, Proc. of 2nd Int'l Conf. on 3D Vision:Tokyo, 517–524, 2014.

5 Á. Santamaria-Navarro and J. Andrade-Cetto, Uncalibrated image-based visual servoing, Proc. of 2013 IEEE Int'l Conf. on Robotics and Automation, Karlsruhe, 5247–5252, 2013.

6 Á. Santamaria-Navarro, V. Lippiello and J. Andrade-Cetto, Task priority control for aerial manipulation, Proceedings of 2014 IEEE Int'l Symp. on Safety, Security and Rescue Robotics, Toyako-cho, 1–6, 2014.

7 Á. Santamaria-Navarro, J. Solá and J. Andrade-Cetto, High-frequency MAV state estimation using low-cost inertial and optical flow measurement units, Proc. of 2015 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems:Hamburg, 1864–1871, 2015.

8 E. H. Teniente and J. Andrade-Cetto, HRA*: Hybrid randomized path planning for complex 3D environments, Proc. of 2013 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, Tokyo, 1766–1771, 2013.

9 R. Valencia, M. Morta, J. Andrade-Cetto and J. M. Porta. Planning reliable paths with pose SLAM, IEEE Transactions on Robotics, 29(4), 1050-1059, 2013.

10 R. Valencia, J. Saarinen, H. Andreasson, J. Vallvè, J. Andrade-Cetto and A. Llilienthal, Localization in highly dynamic environments using dual-timescale NDT-MCL, Proc. of 2014 IEEE Int'l Conf. on Robotics and Automation, Hong Kong, 956–3962, 2014.

11 J. Vallvè and J. Andrade-Cetto, Dense entropy decrease estimation for mobile robot exploration, Proceedings of 2014 IEEE Int'l Conf. on Robotics and Automation, Hong Kong, 6083-6089, 2014.

12 J. Vallvè and J. Andrade-Cetto, Potential information fields for mobile robot exploration, Robotics and Autonomous Systems, 69, 68–79, 2015.

13 J. Vallvè and J. Andrade-Cetto, Active Pose SLAM with RRT*, Proc. of 2015 IEEE Int'l Conf. on Robotics and Automation, Seattle, 2167–2173, 2015.

14 M. Villamizar, F. Moreno-Noguer, J. Andrade-Cetto and A. Sanfeliu, Efficient rotation invariant object detection using boosted random Ferns, Proc. of 2010 IEEE CS Conf. on Computer Vision and Pattern Recognition:San Francisco, 1038–1045, 2010.

15 M. Villamizar, J. Andrade-Cetto, A. Sanfeliu and F. Moreno-Noguer, Bootstrapping boosted random Ferns for discriminative and efficient object classification, Pattern Recognition, 45(9): 3141–3153, 2012.

4.3 Vehicular Instrumentation for the Study of Driver Intent and Related Applications

Steven S. Beauchemin (University of Western Ontario – London, CA)

In this contribution we describe a vehicular instrumentation for the study of driver intent. Our instrumented vehicle is capable of recording the 3D gaze of the driver and relating it to the frontal depth map obtained with a stereo system in real-time, including the sum of vehicular parameters actuator motion, speed, and other relevant driving parameters. Additionally, we describe other real-time algorithms that are implemented in the vehicle, such as a frontal vehicle recognition system, a free lane space estimation method, and a GPS position-correcting technique using lane recognition as land marks.

4.4 Recursive Joint Estimation of Dense Scene Structure and Camera Motion in an Automotive Scenario

Florian Becker (Sony – Stuttgart, DE)

The optical flow induced by a camera moving through a static 3d scene contains valuable information on the geometry. We present an approach to jointly estimating camera motion and a depth map from real-life monocular image sequences which parametrize a flow field. Temporal consistency is exploited in a recursive manner which allows to reduce the estimation task to a series of two-frame problems complemented by an additional temporal smoothness prior. Results for image sequences recorded in a real world traffic scenario are presented.

4.5 Semantic RGB-D Perception for Cognitive Robots

Sven Behnke (Universität Bonn, DE)

Cognitive robots need to understand their surroundings not only in terms of geometry, but they also need to categorize surfaces, detect objects, estimate their pose, etc. In the talk, I will report on efficient methods to address these tasks, which are based on RGB-D sensors. We learn semantic segmentation using random forests and aggregate the surface category in 3D by RGB-D SLAM. We use deep learning methods to categorize surfaces, to recognize objects and to estimate their pose. Efficient RGB-D registration methods are the basis for the manipulation of known objects. They have been extended to non-rigid registration, which allows for transferring manipulation skills to novel objects.

4.6 Second-Order Recursive Filtering on the Rigid-Motion Group SE(3) Based on Nonlinear Observations from Monocular Videos

Johannes Berger (Universität Heidelberg, DE)

Joint camera motion and depth map estimation from observed scene features is a key task in order to reconstruct 3D scene structure using low-cost monocular video sensors. Due to the nonlinear measurement equations that connect ego-motion with the high-dimensional depth map and optical flow, the task of stochastic state-space filtering is intractable.

After introducing the overall problem, the talk focuses on a novel second-order minimum energy approximation that exploits the geometry of $SE(3)$ and recursively estimates the state based on a higher-order kinematic model and the nonlinear measurements. Experimental results for synthetic and real sequences (e.g. KITTI benchmark) demonstrate that our approach achieves the accuracy of modern visual odometry methods.

4.7 Robust Coupling of Perception to Actuation in Dynamic Environments

Darius Burschka (TU München, DE)

I will present methods to represent the state of dynamic environments at a level that is least sensitive to errors in calibration parameters of the sensors. The method is used for a fail-safe implementation of instinctive behaviours on vehicles, like obstacle avoidance. The presented method allows a monitoring of large areas around the vehicle, where a Cartesian representation is not appropriate due to the with distance increasing error in the reconstruction of the three-dimensional information. I will present also the first implementations of this system on a car.

4.8 Direct and Dense Methods for 3D Reconstruction and Visual SLAM

Daniel Cremers (TU München, DE)

Joint work of Cremers, Daniel; Engel, Jakob; Kerl, Christian; Oswald, Martin R; Stuckler, Jorg; Sturm, Jürgen; Stuhmer, Jan; Usenko, Vladyslav

The reconstruction of the 3D world from images is among the central challenges in computer vision. Having started in the 2000s, researchers have pioneered algorithms which can reconstruct camera motion and sparse feature-points in real-time. In my talk, I will present spatially dense methods for camera tracking and reconstruction. They do not require feature point estimation, they exploit all available input data and they recover dense geometry rather than sparse point clouds.

4.9 Pose Estimation and 3D Segmentation using 3D Knowledge in Dynamic Environments

Cédric Demonceaux (University of Bourgogne, FR)

When 2D and 3D cameras observe the same scene, their measurements are usually complementary to each other for scene reconstruction and understanding. A classic example includes 2D cameras capturing high quality texture information and 3D cameras providing accurate location of the scene points. Fusing these complementary measurements has many potential applications such as change detection, scene gaps filling, camera pose correction, and visual odometry. In this talk, we show that the 3D localization of 2D cameras can be improved knowing 3D structure of the scene. Thus, we develop two methods using 3D information on the scene for camera pose estimation. The first one doesn't require 3D feature extraction and doesn't need any geometric hypothesis but it converges in a local optimum. The second one supposes that the scene is composed of planar patches and converges to the global optimum. Then, this 3D information will be used conjointly with 2D images for extracting and reconstructing the background and the dynamic objects of the scene.

4.10 Learning to Drive

Michael Felsberg (Linköping University, SE)

Driving a car is a prototypical example for graded autonomy, where the human driver and the assistance system co-operate. There are various legal, technological, and practical reasons why the human driver is kept in the loop and should be able to override the system's decisions. However, the co-operation, and thus graded autonomy, should be more than just taking over power of command. It is desirable that the assistance system seamlessly acquires new capabilities during the driver's manual intervention, in order to increase the level of autonomy in subsequent operation. Thus, the task for the system is to use input, or precepts, as observed by the human user and output, or actions, as provided by the human user to extend the systems capabilities on the fly. We address this task in the present work and propose a novel approach to online perception-action learning, which lifts human-machine interaction clearly beyond current methods based on switching command. We evaluate our approach on the problem of road following with a model RC-car on reconfigurable tracks indoors, outdoors, and at night. The system's capabilities are continuously extended by interchangeably applying supervised learning (by demonstration), instantaneous reinforcement learning, and unsupervised learning (self-reinforcement learning). The resulting autonomous capabilities go beyond those of existing methods and of the human driver with respect to speed, accuracy, and functionality.

4.11 Drone Vision – Computer Vision Algorithms for Drones

Friedrich Fraundorfer (TU Graz, AT)

Drone Vision – Computer Drones are small scale flying robots and it is predicted that the drone market will see a major growth in the near future. Computer vision will play a major role in controlling and developing autonomous drones. My proposal is to utilize tight IMU-vision coupling for ego-motion estimation of drones. This will result in a new class of fundamental algorithms for ego-motion estimation, being more robust and lots faster. IMU measurements can be used to transform the complex estimation problems into a simpler formulation of vision algorithms for drones.

4.12 Structured Indoor Modeling and/or Uncanny Valley for 3D Reconstruction

Yasutaka Furukawa (Washington University – St. Louis, US)

Depending on how our discussion will go, I will give a talk on one of the following two topics or a mix.

Structured indoor reconstruction. We propose a novel indoor scene reconstruction algorithm. The approach produces a structured 3D model in a top-down manner. The reconstruction algorithm is driven by a indoor structure grammar. The new model representation enables many new applications such as novel indoor scene visualization, inverse CAD, floorplan generation, and tunable reconstruction.

Uncanny Valley for 3D Reconstruction. Accurate 3D reconstruction is usually the key to high quality visualization applications. However, very often, improving reconstruction accuracy degrades the quality of visualization. This issue is little known to researchers, yet very important in practice.

4.13 Underwater Vision: Robots that "see" beneath the surface

Rafael Garcia (University of Girona, ES)

Using vision underwater is a difficult endeavor due to the transmission properties of the medium. Light is absorbed and scattered by the particles suspended in the water column, producing degraded images with limited range, blurring, low contrast and weak colours, among other effects. Moreover, artificial lighting tends to provide non-uniform illumination and introduces shadows in the scene, generating a motion flow that does not obey the dominant motion of the camera.

However, with the adequate processing pipeline, vision can be a powerful tool for underwater robots to explore the ocean.

In this talk we will explain how computer vision techniques can be adapted to the underwater environment if we understand and deal with the different associated problems. An approach to create accurate three-dimensional textured models of the seafloor will be presented. The method de-hazes the images to improve signal-to-noise ratio, then generates a dense cloud of 3D points and is able to compute a meshed surface being robust to common defects in underwater imaging such as high percentage of outliers (due to light backscatter) and point-cloud noise (due to the blurring of forward scattering).

4.14 Semantic Maps for High Level Robot Navigation

Antonios Gasteratos (Democritus University of Thrace – Xanthi, GR)

In the near future domestic robots should be equipped with the potential of producing meaningful internal retalks of their own environment, allowing them to cope a wide range of real-life tasks. Intense research efforts occur to build cognitive robots able to perceive and understand their surroundings in a human-centred manner. Semantic mapping in mobile robotics can constitute a definite solution for this challenge. The semantic map is an augmented representation of the environment that –supplementary to the geometrical knowledge – encapsulates characteristics compatible with human understanding. It provides several algorithmic opportunities for innovative development of applications that will eventually lead to the human robot interaction. This talk will describe the construction of accurate and consistent semantic maps facilitating adequate robot deployment in domestic environments.

4.15 High-level Knowledge in Low-level Vision

Andreas Geiger (MPI für Intelligente Systeme – Tübingen, DE)

1. Stereo techniques have witnessed tremendous progress over the last decades, yet some aspects of the problem still remain challenging today. Striking examples are reflecting and textureless surfaces which cannot easily be recovered using traditional local regularizers. In this work, we therefore propose to regularize over larger distances using object-category specific disparity proposals (displets) which we sample using inverse graphics techniques based on a sparse disparity estimate and a semantic segmentation of the image. The proposed displets encode the fact that objects of certain categories are not arbitrarily shaped but typically exhibit regular structures. We integrate them as non-local regularizer for the challenging object class “car” into a superpixel based CRF framework and demonstrate its benefits on the KITTI stereo evaluation.
2. This work proposes a novel model and dataset for 3D scene flow estimation with an application to autonomous driving. Taking advantage of the fact that outdoor scenes often decompose into a small number of independently moving objects, we represent each element in the scene by its rigid motion parameters and each superpixel by a 3D plane

as well as an index to the corresponding object. This minimal representation increases robustness and leads to a discrete-continuous CRF where the data term decomposes into pairwise potentials between superpixels and objects. Moreover, our model intrinsically segments the scene into its constituting dynamic components. We demonstrate the performance of our model on existing benchmarks as well as a novel realistic dataset with scene flow ground truth. We obtain this dataset by annotating 400 dynamic scenes from the KITTI raw data collection using detailed 3D CAD models for all vehicles in motion. Our experiments also reveal novel challenges which can't be handled by existing methods.

4.16 Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition

Michal Havlena (ETH Zürich, CH)

Structure-based localization is the task of finding the absolute pose of a given query image w.r.t. a pre-computed 3D model. While this is almost trivial at small scale, special care must be taken as the size of the 3D model grows, because straight-forward descriptor matching becomes ineffective due to the large memory footprint of the model, as well as the strictness of the ratio test in 3D. Recently, several authors have tried to overcome these problems, either by a smart compression of the 3D model or by clever sampling strategies for geometric verification. Here we explore an orthogonal strategy, which uses all the 3D points and standard sampling, but performs feature matching implicitly, by quantization into a fine vocabulary. We show that although this matching is ambiguous and gives rise to 3D hyperpoints when matching each 2D query feature in isolation, a simple voting strategy, which enforces the fact that the selected 3D points shall be co-visible, can reliably find a locally unique 2D-3D point assignment. Experiments on two large-scale datasets demonstrate that our method achieves state-of-the-art performance, while the memory footprint is greatly reduced, since only visual word labels but no 3D point descriptors need to be stored.

4.17 Visual-Inertial Navigation for Mobile Robots

Heiko Hirschmuller (Roboception GmbH – München, DE)

Joint work of Hirschmuller, Heiko; Schmid, Korbinian; Suppa, Michael; Lutz, Philipp; Tomic, Teodor; Mair Elmar

Navigation of mobile robots is still challenging, especially in environments that are not prepared for robotics, like at home or outdoors. At the German Aerospace Center (DLR) we have developed passive stereo-vision based navigation that is supported by an inertial measurement unit (IMU). The ego-motion estimation is precise due to visual odometry, robust due to the IMU and fulfils hard-real time constraints for using it directly in the control loop of robots, like for autonomously flying highly agile quadcopters. Several experiments with rovers and flying systems in mixed indoor/outdoor settings proved the concept.

In the DLR spin-off Roboception GmbH we are going to bring such technology as plug and play device into the marked. The talk covers the main concepts and new developments. One of them is the extension of the Semi-Global Matching method for delivering not just disparity, but also error and confidence values for each pixel. The error is given in disparities and has a value of 0.5 for most of the pixels, but can go up to 2. The confidence is the probability that the true disparity is within a three times error interval around the measured disparity. Thus, the error is seen as the standard deviation of an Gaussian error, while the confidence is the probability that the measured value is not an outlier.

References

1 Heiko Hirschmuller, Korbinian Schmid and Michael Suppa, Computer vision for mobile robot navigation, Proceedings of Photogrammetric Week 2015:Stuttgart, 143–154, 2015.

2 Korbinian Schmid, Philipp Lutz, Teodor Tomic, Elmar Mair and Heiko Hirschmuller, Autonomous vision-based micro air vehicle for indoor and outdoor navigation, Journal of Field Robotics, Special Issue on Low Altitude Flight of UAVs, 31(4), 537–570 2014.

3 Heiko Hirschmuller, Stereo Processing by semi-global matching and mutual information, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2), 328–341, 2008.

4.18 Planetary Robotic Vision Processing for Rover Missions

Ben Huber (Joanneum Research – Graz, AT)

The international community of planetary science and exploration has successfully launched, landed and operated about thirty human and robotic missions to the planets and the Moon. They have collected differing numbers of surface imagery that have only been partially utilized throughout these missions and thereafter for further scientific application purposes. The data for most of these missions including meta-data is publicly available. Many of the mentioned missions rely on stereo imagery for navigation and offer huge datasets that can be reconstructed in 3D and put into a common coordinate context. By doing this we generate datasets with a huge benefit for scientific geological analysis in rendered 3D space based on mission data that has been almost forgotten in planetary data archives.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 312377 PRoViDE.

4.19 Model-based Object and Deformation Tracking with Robust Global Optimization

Reinhard Koch (Universität Kiel, DE)

Model-based analysis, also termed Analysis-by-Synthesis or Analysis from Generative models, is a powerful tool to solve ill-posed problems like 3D object surface tracking from images. A parametric model of the object in focus is generated and the visual appearance and motion of the object is synthesized from the model and compared with the visual input by a cost or fitness function. Adapting the model parameters according to the cost function solves the tracking problem. In order to cope with possibly high-dimensional parameter space, efficient and robust non-local stochastic estimators are needed. In my talk I will outline the AbS principle and discuss the optimizer and applications. Examples are tracking of multiple animals in confined housing, deformation of thin plate models for human user interaction, and others.

4.20 Image based Navigation for Exploration Probe

Takashi Kubota (ISAS/JAXA, JP)

This talk firstly introduces future lunar or planetary exploration plans, which consist of lunar, Mars and asteroid exploration. Then robotics technology is shown for lunar or planetary exploration. Vision system makes important roles in deep space exploration for efficient and safe exploration. This talk presents the intelligent system for navigation, path planning, sampling, etc. Especially image based navigation schemes are presented in detail.

4.21 Stereo Vision for Future Autonomous Space Exploration Robots

Lazaros Nalpantidis (Aalborg University Copenhagen, DK)

Space exploration rovers need to be highly autonomous because the vehicle should spend as much of its traverse time as possible moving, rather than waiting for delayed and often interrupted teleoperation commands. Autonomous behavior can be supported by vision systems that provide wide views to enable navigation and 3D reconstruction, as well as close-up views ensuring safety and providing reliable odometry data.

This talk presents the design, development and testing of such a stereo vision system for a space exploration rover. This system was designed with the intention of being efficient, low-cost, accurate and was ultimately implemented on an FPGA platform. We are discussing our experiences with this system and highlight useful lessons learned.

4.22 Structure and Motion – Challenges and solutions for real time geometric estimation from video

Mikael Persson (Linköping University, SE)

In this talk I introduce the cv4x SfM system, which achieved state of the art results on the challenging KITTI odometry benchmark earlier this year(2015). I motivate the choice in reconstruction method, the bootstrap tracking by matching scheme used and how perceptual aliasing was addressed. Visual odometry systems such as cv4x, in particular if fused with a strong loop closure system, achieve excellent and more importantly sufficient results, but several challenges remain: The trajectory prediction of independently moving objects is of particular interest in autonomous driving and is principally, if not practically, the same problem as IMU-free VO and as such my current focus of research.

4.23 Efficient Block Optimization Methods for Computer Vision

Thomas Pock (TU Graz, AT)

In this talk I will discuss recent advances in block optimization methods for minimizing non-smooth optimization problems in computer vision and image processing. It turns out that a large class of 2D and 3D total-variation regularized problems can be reduced to an algorithm that computes exact solutions with respect to certain subsets of the variables in each iteration. For example, if the subsets are 1D total variation problems, we can efficiently compute their solutions based on dynamic programming. Furthermore, we can make use of gradient acceleration techniques to additionally speed up the algorithms. I will show applications to computing globally optimal minimizers of total variation regularized stereo problems.

4.24 Toward Highly Intelligent Automobiles

Danil V. Prokhorov (Toyota Research Institute North America – Ann Arbor, US)

Intelligent automobiles a.k.a. self-driving, autonomous or highly automated cars are capturing people's imagination while opening up new opportunities for research in many areas including robotics, machine learning and vision. In my talk I overview the state of art in highly automated cars and discuss an example of near-production AHDA car of Toyota, as well as my personal experience with ACC-equipped vehicles. Then I discuss an example of on-going research project of a car capable of making a variety of autonomous decisions on public roads, with the focus on the roundabout maneuver. This maneuver illustrates an importance of the holistic perception-action system approach, rather than module-by-module considerations still prevalent in this field. In conclusion I offer my view of open challenges for intelligent automobiles.

4.25 Efficient Multi-view Semantic Segmentation

Hayko Riemenschneider(ETH Zürich, CH)

There is an increasing interest in semantically annotated 3D models, e.g. of cities. The typical approaches start with the slow semantic labelling of all the images used for the 3D model. The inherent redundancy among the overlapping images calls for more efficient solutions. This work deals with two an alternative approaches. First, we exploit the geometry of a 3D mesh model to predict the best view before the actual labelling. For this we find the single image part that bests supports the correct semantic labelling of each face of the underlying 3D mesh. Second, we directly use the 3D point cloud itself, skipping any image processing entirely. This pure 3D approach relies solely on 3D surface features (and a bit of classic RGB) and provides state-of-the-art results without any heavy 2D image features. In both works we show how to significantly speedup the semantic segmentation and even increase the accuracy – leaving the question – how much of 2D images is needed for 3D semantic segmentation?

4.26 The Limits of Pose Estimation in Very Large Maps

Torsten Sattler (ETH Zürich, CH)

In many applications of autonomous vehicles, we can safely assume that there exists a 3D map of the scene the vehicle operates in, which can be used for navigation and localization. One major problem of maps covering a very large area is that they contain many structures with globally repeating appearance, causing problems when trying to match structures between a query image and the map. Common large-scale localization approaches operate under the assumption that we can recover the pose of the query image as long as we find enough good matches and as long as the pose estimation process is robust enough to large quantities of wrong matches. Unfortunately, current pose estimation strategies have problems dealing with too many matches. In this talk, I will discuss a truly scalable pose estimation strategy. Using this strategy, we will show that there are limits to the approach of just using more and more matches and hoping that pose estimation will be able to recover the correct pose.

4.27 From Frames to Events: Vision for High-speed Robotics

Davide Scaramuzza (Universität Zürich, CH)

Autonomous micro drones will soon play a major role in search-and-rescue and remote-inspection missions, where a fast response is crucial. They can navigate quickly through unstructured environments, enter and exit buildings through narrow gaps, and fly through collapsed buildings. However, their speed and maneuverability are still far from those of birds. Indeed, agile navigation through unknown, indoor environments poses a number of challenges for robotics research in terms of perception, state estimation, planning, and control. In this talk, I will give an overview of my research activities on visual inertial navigation of quadrotors, from slow navigation (using standard frame-based cameras) to agile flight (using event-based cameras).

4.28 Towards 3D Scene Understanding

Bernt Schiele (MPI für Informatik – Saarbrücken, DE)

Inspired by the ability of humans to interpret and understand 3D scenes nearly effortlessly, the problem of 3D scene understanding has long been advocated as the "holy grail" of computer vision. In the early days this problem was addressed in a bottom-up fashion without enabling satisfactory or reliable results for scenes of realistic complexity. In recent years there has been considerable progress on many sub-problems of the overall 3D scene understanding problem. As the performance for these sub-tasks starts to achieve remarkable performance levels we argue that the problem to automatically infer and understand 3D scenes should be addressed again.

This talk highlights recent progress on some essential components (such as object recognition and person detection), on our attempt towards 3D scene understanding, as well as on our work towards activity recognition and the ability to describe video content with natural language. These efforts are part of a longer-term agenda towards visual scene understanding. While visual scene understanding has long been advocated as the “holy grail” of computer vision, we believe it is time to address this challenge again, based on the progress in recent years.

4.29 Tracking and Mapping in Project Tango

Jürgen Sturm (Google – München, DE)

Google’s Project Tango aims to provide a mobile solution for visual-inertial 6-DOF motion estimation and dense 3D reconstruction. In my talk, I will give a technical presentation of the algorithms underlying the Tango API, including visual-inertial odometry, SLAM, loop closure detection, re-localization and 3D reconstruction. During my talk, I will present several live demos on a Tango tablet.

4.30 Deep Learning for Visual Place Recognition and Online 3D Reconstruction

Niko Sünderhauf (Queensland University of Technology – Brisbane, AU)

In the first part of this talk I will summarize our recent work on visual place recognition in changing environments using deep convolutional network features.

In the second part I talk about some lessons learned when applying ConvNets in robotics (e.g. for object detection on a mobile robot) and the gaps between the computer vision community and robotics in that particular area. I hope to induce a discussion on what we as a community can do to bridge this gap.

4.31 Large-scale Visual Place Recognition – Current challenges

Akihiko Torii (Tokyo Institute of Technology, JP)

Large-scale visual place recognition (VPR) takes an important role for localization of robots and autonomous cars, e.g. rough initial localization. In this seminar, we first compare key properties of compact image descriptors – Bag of Visual Words (BoVW) and Vector of Locally Aggregated Descriptors (VLAD) – popularly used in VPR. On top of the decent analysis of these image retalks, we discuss challenges in VPR, e.g. repetition, illumination changes, change of seasons, and aging that give major appearance changes among testing-query and database images. We show that the adaptive soft-assignment scheme on BoVW is effective on the street-level visual place recognition. We also show dense feature detection followed by VLAD representation gives a significant improvement in localization performance and expanding the database by view synthesis gives an additional gain on the challenging datasets.

4.32 3D Scene Understanding for Autonomous Driving

Raquel Urtasun (University of Toronto, CA)

Developing autonomous systems that are able to assist humans in everyday's tasks is one of the grand challenges in modern computer science. Notable examples are personal robotics for the elderly and people with disabilities, as well as autonomous driving systems which can help decrease fatalities caused by traffic accidents. In order to perform tasks such as navigation, recognition and manipulation of objects, these systems should be able to efficiently extract 3D knowledge of their environment. In this talk, I'll show how graphical models provide a great mathematical formalism to extract this knowledge. In particular, I'll focus on a few examples, including 3D reconstruction, 3D object and layout estimation and self-localization.

4.33 Direct SLAM Techniques for Vehicle Localization and Autonomous Navigation

Vladyslav Usenko (TU München, DE)

Joint work of Cremers, Daniel; Engel, Jakob; Jorg Stuckler, Jorg; von Stumberg, Lukas; Usenko, Vladyslav

Localization and mapping are two very important challenges for autonomous vehicles. Even though many different types of sensors can be used for this purpose, camera based solutions gain popularity because of the low costs, small weight and simple mechanical design. In my talk I present several extensions to the LSD-SLAM – camera based large-scale direct semi-dense slam method, that enable reliable operation on the real-world data from the vehicles. In particular, I present an extension of the method to the stereo-camera setup and tight integration with Inertial Measurement Unit, and demonstrate an autonomous exploration and control on a consumer grade flying robot.

4.34 Learning See in a Virtual World

David Vázquez Bermudez (Autonomous University of Barcelona, ES)

The ADAS group from the Computer Vision Center based at the Universitat Autònoma de Barcelona, has an extensive experience developing ADAS systems such as Lane Departure Warning, Collision Warning, Automatic Cruise Control, Pedestrian Protection, Headlights Control, etc. Currently ADAS is developing an Autonomous Vehicle based on relatively cheap sensors such cameras, IMU and GPS. In this talk we will give a short overview of the ADAS systems developed until now by the group and the Autonomous Vehicle project. Then we will explain in more detail two parts of the autonomous vehicle that has been awarded by the IEEE Intelligent Transportation Systems Society Spanish Chapter. The use of virtual images to training models that are able to operate in a real world (Best Ph.D. Thesis award) and a vehicle localization system based on GPS, IMU and cameras (Accesit M.Sc award).

4.35 Realizing Self-Driving Car

Andres Wendel (Google Inc. – Mountain View, US)

Self-driving vehicles are coming. They will save lives, save time and offer mobility to those who otherwise don't have it. Eventually they will reshape the world we live in. A dedicated team at Google has spent the last few years moving self-driving vehicles closer to reality. New algorithms, increased processing power, innovative sensors and massive amounts of data enable our vehicles to see further, understand more and handle a wide variety of challenging driving scenarios. Our vehicles have driven over a million miles on highways, suburban and urban streets. Through this journey, we've learned a lot; not just about how to drive, but about interacting with drivers, users and others on the road, and about what it takes to bring an incredibly complex system to fruition. In my talk, I share some insights in how the technology works, how we have rolled out our new prototype vehicles to public roads, and which edge case situations we have to solve.

5 Working groups

5.1 Sensing

Editor: Andrés Bruhn
Topic: Low-Level Sensing

5.1.1 Workgroup members in alphabetical order

Andrés Bruhn	(Universität Stuttgart, DE)
Florian Becker	(Sony – Stuttgart, DE)
Johannes Berger	(Universität Heidelberg, DE)
Darius Burschka	(TU München, DE)
Ben Huber	(Joanneum Research-Graz, AT)
Reinhard Koch	(Universität Kiel, DE)
Lazaros Nalpantidis	(Aalborg University Copenhagen, DK)
Thomas Pock	(TU Graz, AT)

5.1.2 Discussion Summary

This working group discussed aspects of low-level sensing methods for autonomous vehicles and probes. While most systems for autonomous driving rely on the same types of modules – e.g. algorithms for motion estimation, stereo reconstruction, and scene flow computation – those modules are typically designed and evaluated separately from the remaining system. Evidently, this makes it difficult to integrate feedback in terms of scene understanding, which would be likely to improve the robustness of such algorithms in difficult situations, i.e. under adverse weather conditions. Moreover, the learning of suitable models or model components for specific scenarios is becoming increasingly important as the integration of previously learned priors may improve the quality of the algorithms as well. Also from an evaluation viewpoint, there is a clear need for improvement. Currently there is a clear lack of suitable benchmarks to evaluate the quality of vision algorithms for autonomous driving. While there are at least some benchmarks that can be used to evaluate the performance of low-level algorithms separately (e.g. the KITTI Benchmark Suite), it remains unclear which accuracy and robustness demands complex systems for autonomous driving actually have w.r.t. to the performance of their underlying modules. Hence it is hardly possible to predict the performance and robustness of such methods for real applications such as autonomous driving. Finally, the use of different hardware for image acquisition may significantly improve both performance and speed of low-level algorithms. One the one hand, one may consider the use of high speed cameras to avoid ambiguous large displacements which still pose a problem for most applications. On the other hand, it may be worthwhile to investigate the usefulness of "differential" cameras that allow an reduction of the processed data by only providing information in terms of image changes. In detail, the following research questions have been discussed:

How can low-level models be further improved?

- flexible system design (modules) vs. robustness (high level knowledge).
- however: jointly solving strongly related tasks may improve performance.
- need of joint modelling and inference in terms of holistic approaches.
- moreover: learning of model components based on given application.

How can benchmarking be improved towards practical relevance?

- in practice: absolute accuracy not that important.
- algorithm must be "sufficiently accurate" for a certain application.
- evaluation as part of the entire vision system.
- robustness matters: determine breaking point under certain degradations.

How can the image acquisition process be improved?

- in general: higher frame rates desirable for motion estimation.
- simpler algorithms sufficient → faster computation.
- less complex motion, smaller displacements → higher accuracy.

What are suitable representations when extracting information?

- typically: not all pixels needed for making decisions.
- only consider locations that deviate from expected behaviour (e.g. intensity changes, deviations from high-level models).
- "differential" cameras (e.g. event cameras).

5.2 Mapping for Autonomous Vehicles and Probes

Editor: Hayko Riemenschneider
Topic: Offline Mapping

5.2.1 Workgroup members in alphabetical order

Yasutaka Furukawa	(Washington University St. Louis, US)
Antonios Gasteratos	(Democritus University of Thrace – Xanthi, GR)
Michal Havlena	(ETH Zürich, CH)
Hayko Riemenschneider	(ETH , Zürich, CH)
Torsten Sattler	(ETH Zürich, CH)
Akihiko Torii	(Tokyo Institute of Technology, JP)
Vladyslav Usenko	(TU München, DE)

5.2.2 Discussion Summary

The what, where, when, how, and who of mapping. This working group defined mapping as the process to create (offline/online) environment maps including road an urban environment, lane markings and traffic symbols as well as dynamic obstacles like pedestrians or weather conditions. One main topic is the distinction between offline prior mapping and online mapping. The group concluded that the hard cases, those which currently pose the most challenges (dynamic objects and up to date information), can only be solved in online mapping whereas offline mapping can provide a solid environment yet by definition will always be out of date. Hence, the question arises of the use cases for offline mapping, e.g. route navigation planning.

How to technically create environment maps?

- offline mapping will benefits from the richness of all sensors (vision, LIDAR, etc).
- online mapping also, yet for real time purposes needs specializations (instant LIDAR results vs stixel like abstractions).
- transfer manual annotation from 2D to 3D or vice versa, needed for guarantee on quality.
- define levels of details for roads, surroundings and buildings.
- formalism of maps, structures, relationships of contents in there (roads are connecting).

When to create environment maps?

- temporally changing maps, need for continuously updating.
- integrating visual information acquired by different companies, people, cars, ...
- collecting data, by own cars, taxi, trams, community services.
- how fast should be the update vs on the drive will always be fastest.
- how to integrate multimodal data coming apart from visual data.
- long term changes, building construction.
- short term changes, e.g. parking cars, construction sites, people movement updating maps give more useful prior information for autonomous driving/routing, e.g. once construction signs found, we have no need to drive there.
- other important issues: security, redundancy, fall back of the map creation.

Who is responsible for the creation of maps?

- service levels agreement for quality.
- will there be a uniform/standard format of global map?
- consortium of car companies and users to define these standards no common maps since business model. coverage of the countries in terms of where are maps needed and to what detail.
- accuracy of the maps w.r.t. coverage, not everywhere is a cm/time accuracy needed.
- crowdsourcing for every car.
- human control and verification of structural changes (suggested by vision, LIDAR).

Where and what is included in the environment maps?

- we should know limitations of online/offline mapping!
- what is the difference between online/offline mapping?
- online: annotation, change detection → impossible to do online. only the surrounding areas.
- only does simple dynamic obstacles/events detection (no understanding what it is).
- what if non standard events happen. if the roads are covered by bus. deadlock situation.
- offline: semantics: soft not binary decision: continuous occlusion space and object classes.
- classifiers on actions/intent of others, to allow high level interpretation when breaking rules.

5.3 Beyond Deep Learning

Editor: Michael Felsberg
Topic: Deep Learning

5.3.1 Workgroup members in alphabetical order

Andreas Geiger	(MPI für Intelligente Systeme – Tübingen, DE)
Atsushi Imiya	(Chiba University, JP)
Bernt Schiele	(MPI für Informatik – Saarbrücken, DE)
José M. Alvarez	(NICTA – Canberra, AU)
Jürgen Sturm	(Google – München, DE)
Michael Felsberg	(Linköping University, SE)
Niko Sünderhauf	(Queensland University of Technology – Brisbane, AU)
Rafael Garcia	(University of Girona, ES)
Raquel Urtasun	(University of Toronto, CA)
Sven Behnke	(Universität Bonn, DE)

5.3.2 Discussion Summary

Recurrent and Dynamic Networks with Structural Models. Future work will have to address procedural fundamentals of the learning algorithm and the network:

- How to realize deep learning of recurrent networks?
- How to realize networks that learn layered dynamic processes?
- How to regularize with known structural and geometrical models?
- How to enforce invariance beyond shift invariance?
- How to inject hard constraints?

These issues establish an engineering – understanding trade-off. Advanced visualizations and modelling of solution manifolds are required for future progress.

Learning Process and Training Data. It has been reflected that previous research often evaluated sub-tasks, such as detection or recognition, instead of system-level performance. The latter will, in most cases, require embodiment and thus perception-action learning. Future challenges will include:

- How to perform reinforcement learning on deep networks?
- How to generate training data with sufficient volume and quality?
- How to synthesize and augment data?
- How to regularize learning with known stochastic models to avoid overfitting?
- How to assess performance on system-level tasks in relation to other techniques such as random forests?

These issues establish a major challenge on the empirical analysis of deep learning. An evaluation methodology has to be developed to assess progress properly.

Participants

- José M. Alvarez
NICTA – Canberra, AU
- Juan Andrade-Cetto
UPC – Barcelona, ES
- Steven S. Beauchemin
University of Western Ontario – London, CA
- Florian Becker
Sony – Stuttgart, DE
- Sven Behnke
Universität Bonn, DE
- Johannes Berger
Universität Heidelberg, DE
- Andrés Bruhn
Universität Stuttgart, DE
- Darius Burschka
TU München, DE
- Daniel Cremers
TU München, DE
- Krzysztof Czarnecki
University of Waterloo, CA
- Cédric Demonceaux
University of Bourgogne, FR
- Michael Felsberg
Linköping University, SE
- Friedrich Fraundorfer
TU Graz, AT
- Yasutaka Furukawa
Washington University – St. Louis, US
- Rafael Garcia
University of Girona, ES
- Antonios Gasteratos
Democritus Univ. of Thrace – Xanthi, GR
- Andreas Geiger
MPI für Intelligente Systeme – Tübingen, DE
- Michal Havlena
ETH Zürich, CH
- Heiko Hirschmuller
Roboception GmbH – München, DE
- Ben Huber
Joanneum Research – Graz, AT
- Atsushi Imiya
Chiba University, JP
- Reinhard Koch
Universität Kiel, DE
- Takashi Kubota
ISAS/JAXA – Sagamihara, JP
- Lazaros Nalpantidis
Aalborg Univ. Copenhagen, DK
- Mikael Persson
Linköping University, SE
- Thomas Pock
TU Graz, AT
- Danil V. Prokhorov
Toyota Research Institute North America – Ann Arbor, US
- Sebastian Ramos
Daimler AG-Boblingen, DE
- Hayko Riemenschneider
ETH – Zürich, CH
- Torsten Sattler
ETH Zürich, CH
- Davide Scaramuzza
Universität Zürich, CH
- Bernt Schiele
MPI für Informatik – Saarbrücken, DE
- Jürgen Sturm
Google – München, DE
- Niko Sünderhauf
Queensland University of Technology – Brisbane, AU
- Akihiko Torii
Tokyo Institute of Technology, JP
- Raquel Urtasun
University of Toronto, CA
- Vladyslav Usenko
TU München, DE
- David Vázquez Bermudez
Autonomus University of Barcelona, ES
- Andreas Wendel
Google Inc. – Mountain View, US
- Christian Winkens
Universitat Koblenz-Landau, DE

Report of Dagstuhl Seminar 15462

The Mobile Revolution – Machine Intelligence for Autonomous Vehicles

Edited by

Wolfram Burgard[1], Uwe Franke[2], Markus Enzweiler[3], and Mohan Trivedi[4]

1 **Universität Freiburg, DE,** burgard@informatik.uni-freiburg.de
2 **Daimler AG – Sindelfingen, DE,** uwe.franke@daimler.com
3 **Daimler AG – Böblingen, DE,** markus.enzweiler@daimler.com
4 **University of California, San Diego – La Jolla, US,** mtrivedi@ucsd.edu

Abstract

This report documents the Dagstuhl Seminar 15462 "The Mobile Revolution – Machine Intelligence for Autonomous Vehicles". The seminar has discussed the state-of-the-art and provided a consistent vision on the topic of intelligent autonomous vehicles. It has served as a communication platform between the various sub-communities involved, by bringing together key persons in industry and academia in their respective fields. Additionally, relations between different disciplines of intelligent transportation systems have be identified and exploited. The seminar has allowed its participants to bridge the gap between foundational research and real-world applications by identifying further research directions and initiating interdisciplinary collaborations.

Seminar November 10–13, 2015 – http://www.dagstuhl.de/15462
1998 ACM Subject Classification Autonomous vehicles, Robotics, Vision and Scene Understanding
Keywords and phrases autonomous driving, digital maps, environment perception, machine intelligence, robotics, situation awareness, traffic safety
Digital Object Identifier 10.4230/DagRep.5.11.62

1 Executive Summary

Markus Enzweiler

Motivation and Perspective

Machine intelligence, robotics and computer vision, formerly rather peripheral disciplines of computer science, are in fact already with us today and have a familiar embodiment – the modern vehicle. Systems that are currently available strongly couple interdisciplinary fundamental research with complex practical realizations. The vision of autonomous vehicles in particular has a surprisingly long history with first prototypical implementations going back to the early 1980s. What started then as a dream of pioneers such as Ernst Dickmanns is actually happening right now – we are on the verge of a mobile revolution with self-driving vehicles as its central foundation. The tremendous progress made in the last years has been sparked by the increased methodical and technically availability of better sensors, sophisticated algorithms, faster computers and more data.

The Mobile Revolution – Machine Intelligence for Autonomous Vehicles, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 62–70
Editors: Wolfram Burgard, Uwe Franke, Markus Enzweiler, and Mohan Trivedi
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

But, we are not quite there yet. Autonomous systems make extreme demands on system performance, quality, availability, reliability and verification that significantly increase with the rising degree of automation. Such diverse requirements give rise to numerous problems and open questions that are currently addressed in substantial academic and industrial research activities in many fields of computer science and engineering. Extraordinarily positive innovation effects result from the knowledge transfer between industry and academia, as successfully demonstrated by initiatives such as Uni-DAS or DRIVE-U. The increasing relevance and interest in the computer science community, particularly in the fields of robotics, computer vision and machine learning is evident through an abundance of papers and workshops at major computer science conferences.

This seminar has brought together the leading experts from both academia and industry to discuss the state-of-the-art, identify further research directions and refine the overall vision of intelligent autonomous vehicles into a consistent and practicable picture.

Seminar Topics and Structure

The ultimate design goal for autonomous systems is to mimic human behavior in terms of understanding and effortlessly acting within a dynamic human-inhabited environment. Although artificial sensors emulating the human sensory systems are nowadays widely available, current autonomous systems are still far behind humans in terms of understanding and acting in real-world environments. The chief reason is the (theoretical and practical) unavailability of methods to reliably perform perception, recognition, understanding and action on a broad scale, i.e. not limited to isolated problems.

Following the classical perception-action cycle, the central topics of the seminar have evolved around four key questions posed from the perspective of an autonomous vehicle:

- What do I perceive and how can I interpret this?
- Where am I and what do I do next?
- How can I build up experience and learn?
- Am I capable of this task?

More specifically, the seminar has stimulated research and discussions through several talks on the following topics:

- Intelligent Robotics
- Digital Maps
- Human-centered Intelligent Vehicles
- Verification and Validation
- Limitations and Perspectives

The seminar was held in a very interactive workshop style allowing for ample time for thorough discussions. There were four main sessions with talks and discussions, c.f. the seminar schedule in Section 4, focusing on autonomous driving projects, mapping and localization, sensing, as well as evaluation and approval. The first session on state-of-the-art autonomous driving projects has been co-organized with Seminar 15461 as a joint session.

2 Table of Contents

3 Overview of Talks

3.1 Realizing Self-Driving Cars

Andreas Wendel (Google X, Mountain View, US)

Self-driving vehicles are coming. They will save lives, save time and offer mobility to those who otherwise don't have it. Eventually they will reshape the world we live in. A dedicated team at Google has spent the last few years moving self-driving vehicles closer to reality. New algorithms, increased processing power, innovative sensors and massive amounts of data enable our vehicles to see further, understand more and handle a wide variety of challenging driving scenarios. Our vehicles have driven over a million miles on highways, suburban and urban streets. Through this journey, we've learned a lot; not just about how to drive, but about interacting with drivers, users and others on the road, and about what it takes to bring an incredibly complex system to fruition. In my talk, I share some insights in how the technology works, how we have rolled out our new prototype vehicles to public roads, and which edge case situations we have to solve.

3.2 Autonomous Vehicles – Relations to Human Intelligence

Klaus Bengler (TU München, DE)

Human behavior and change of human behavior is basis not only for critical incidents and accidents but also for efficient traffic and mitigation. The introduction of automation is going to replace or even should support these mechanisms. HMI design plays a dominant role here. Furthermore the introduction of automation will lead to behavioral changes regarding users of the automation and other traffic participants. These effects need further research.

3.3 The Use of 3D Prior Maps in Automated Driving

Ryan Eustice (University of Michigan – Ann Arbor, US)

Self-driving test vehicles have become a reality on roadways and there is an ever present push toward making them a consumer product in the not so distant future. In this talk, I will give an overview of some of our on-going work (in collaboration with Ford Motor Company) in full-scale automated driving. In particular, we'll look at some of our successes in high definition map building and precision localization, including our recent work in cross-modality localization using vision within a priori LIDAR maps. We'll also reflect upon the challenges ahead in perception and human factors.

3.4 HERE Vision of the Future for Autonomous Vehicles

D. Scott Williamson (HERE – Chicago, US) and Alex Goldberg (HERE – Carlsbad, US)

At HERE we are trying to answer three questions: “Where exactly am I?”, “What Lies Ahead?”, and “How can I get there comfortably?”. Maps are needed for autonomous driving as an extended sensor of the road ahead in order to make autonomous driving safer and more comfortable. Comfort is enabled by permitting planning beyond sensor range, and providing information about historical driving behaviour. We present our vision and strategies to address these challenges. Our vision includes a map with accurate lane level geometry, localization features, live traffic events, and speed profiles. This information is delivered via a cloud service that is always enabled providing the freshest data available. In order to achieve the freshness required for autonomous driving we see great potential in utilizing sensor feedback from autonomous vehicles. We see a need for industry alignment on features required for localization, this requires active area of industry collaboration.

3.5 How to Address the Approval Trap for Autonomous Vehicles

Hermann Winner (TU Darmstadt, DE), Maren Graupner (TU Darmstadt, DE), and Walther Wachenfeld (TU Darmstadt, DE)

Autonomous vehicles will pass the technology readiness level of prototype demonstrators in an operational environment soon. Thereby the human ability to control the vehicle must be fully replaced by a technical system. Such a cognitive system that perceives and processes the complex world in public traffic has never before been approved for series production. The so called approval trap appears which means that a ready to use developed autonomous vehicle cannot be released due to the lack of safety validation concepts. What could be the way out? Two approaches with different introduction strategies will lead to gain the needed knowledge for validation. For both efficient test tools and systematical reduction of test cases have to be developed. A method of decomposition of the entire automation process into functional layer corresponding to human driving skills is introduced. This decomposition opens an orthogonal way for testing machine driving efficiently. But, in any case the safety validation has to be part of the first introduction of autonomous driving by risk limitation.

3.6 Driving in Unstructured Environments

Hans-Joachim Wünsche (Universität der Bundeswehr – München, DE)

I started my presentation with my definition of “unstructured environment”: hard to “see´´ road/paths (view through on-board camera onto a muddy forest road during Elrob 2007), poor maps and very poor GPS (from Elrob 2009). This motivates my research: concentrate on perception (even it is very tough), forget precise metric maps (rather create topological maps containing relevant, and observable /re-cognizable landmarks and try to get by without GPS. Examples of perception shown in several videos include detection of forest roads and unknown off-road intersections by vision and by raw data and feature based fusion of vision, lo-light, thermal cameras and Lidar data, both during day time and data in the middle of the night. Raw data fusion is also used to build rich, colored 3D-terrain maps of the immediate vehicle surrounding from Lidar point clouds colored through a color camera yawing left & right and encoding terrain steepness and info (if available) from probably available maps such as vegetation. Landmarks are perceived and have to be so compact as to transmit them via a 9600 baud radio link to other vehicles, which then use this info to build topological maps. These help to navigate autonomously in order to follow each other at larger distances, or to go from A to B without GPS, just localizing relative to landmarks. A video showed successful action, even if the landmarks are seen "from the rear" because the path is driven in reverse order than recorded. The talk closed with emphasizing human like behavior also for "object relational" trajectory planning and recursive situation assessment, to concentrate on what's relevant and less on what's available.

3.7 Human Factors in Intelligent Vehicles

Mohan Trivedi (University of California, San Diego – La Jolla, US)

Designing fully autonomous robotic vehicles which can drive on roads does not require models of drivers and how they interact with vehicles. In contrast – design of humanized "intelligent" vehicles especially those for active safety that prevent accidents, requires understanding of human behavior, modeling of human-vehicle interaction, activities inside the vehicle, and prediction of human intent. We present an overview of a fifteen year old journey of research into such activities and share our findings and experience of basic research as well as road-tested experimental studies in the real-world driving conditions. A holistic framework that utilizes all contextual cues from looking outside as well as inside the vehicle and a sparse Bayesian learning framework has provided convincing results for deployment of predictive, robust and reliable performances.

3.8 Towards Cooperative Autonomous Vehicles

Christoph Stiller (KIT – Karlsruher Institut für Technologie, DE)

In my talk I first revisited the challenges in Automated Driving and outlined the lessons learned from it. While the Urban Challenge 2007 showed that map-based driving with Laser- Scanners and DGPS/INS was feasible in urban areas with low buildings, the Grand Cooperative Driving Challenge in the Netherlands 2011 demonstrated cooperative maneuvers between cars fro, teams that have developed their systems independently for the first time. In collaboration of FZI/KIT with Daimler, the Bertha Benz Route automation showed that autonomous vehicles mainly running with vision sensors may navigate autonomously through a densely populated area in Germany. The talk closed with open issues and remaining challenges that need to be addressed to make level 4/5 automated vehicles a reality.

3.9 Sensors and Perceptual Algorithms

Michael James (Toyota Research Institute North America- Ann Arbor, US)

Highly automated vehicles depend on a wide range of sensors, information sources, and perceptual algorithms to make sense of their surroundings. In our view, this includes direct sensing technologies such as cameras, lidar, and radar; sources of information such as high-definition maps; and sources of remote data such as V2X. We present an overview of how these sensors are used at Toyota Research Institute North America in our highly automated test vehicle, similar in many respects to other research vehicles around the world. Then we present an overview of current and in-development sensors and their capabilities and limitations, as well as a set of potential open questions about how sensors and perception algorithms can be developed in the future.

3.10 Toward Fully Automated Driving

Jan Becker (BOSCH Research Center – Palo Alto, US)

In this talk, we present Bosch's approach to highly and fully automated driving. We define our vision and roadmap and introduce highway pilot as an example for a highly automated system. Then we present underlying technologies, localization and mapping, perception and planning. The talk emphasizes remaining challenges from the current state of the art to series production. Notable challenges are surround sensing robust in all use cases, safety and security i.e. protection against technical failures as well as against deliberate attacks, global standards for legislation and new liability regulations, commercially available precise and up-to-date maps, and new system architecture requirements which includes redundancies for sensing, ECUs, and actuators. We close with an analysis of human factor requirements, specifically regarding transition between automated and manual driving, as well as of the user experience on a highly or fully automated vehicle.

3.11 Coast to Coast and Urban Driving Experience

Serge Lambermont (Delphi Labs – Mountain View, US)

This talk outlines the market development and market drivers for autonomous vehicles. It briefly describes the conventional automotive market with increased ADAS functionality and level 2 traffic jam assist as well as highway pilot. After this adjacent markets as Mobility on Demand are described. Subsequently the sensor impact and new sensor technologies are described. Then, experiences on an automated drive from San Francisco to New York City, as well as urban drives in Las Vegas and Silicon Valley are provided. In the last part, situations and obstacle are described in an occurrence versus severity plane.

3.12 Autonomous Automobiles – Current Challenges in Research and Development

Markus Maurer (TU Braunschweig, DE)

In this talk I summarized the current challenges in the research and development of autonomous automobiles. To make sure everybody had the same understanding of "Autonomous Driving" I started with a short definition of autonomy. I also made my personal perspective transparent: it is basically a systems perspective. In the main part of the talk I addressed the following research fields on the way towards autonomous driving: requirements on infrastructure, system architecture, security, functional safety, perception, decision making, cooperative testing, open systems, acceptance and risk management and education.

4 Schedule

Wednesday, November 11, 2015 (partly shared with seminar 15461)

09:00–09:15	Welcome Address	Organizers
09:15–10:35	Autonomous Driving Projects (I)	Christoph Stiller, Raul Rojas
10:35–10:55	Coffee Break	
10:55–12:15	Autonomous Driving Projects (II)	Markus Maurer, Andreas Wendel
12:15–13:30	Lunch	
13:30–14:30	Introduction Round	All Participants
14:30–15:50	Autonomous Driving Projects (III)	Klaus Dietmayer, Wolfram Burgard
15:50–16:10	Coffee Break	
16:10–17:30	Autonomous Driving Projects (IV)	Hans-Joachim Wünsche, Jan Becker

Thursday, November 12, 2015

09:00–10:00	Session: Mapping and Localization (Chair: Mohan Trivedi)	Alex Goldberg, Scott Williamson, Ryan Eustice
10:00–10:10	Coffee Break	
10:10–12:00	Group Work on Mapping and Localization	
12:00–13:30	Lunch	
13:30–14:30	Session: Sensing (Chair: Uwe Franke)	Serge Lambermont, Michael James
14:30–14:40	Coffee Break	
14:40–17:00	Group Work on Sensing	

Friday, November 13, 2015

09:00–10:00	Session: Evaluation and Approval (Chair: Markus Enzweiler)	Hermann Winner
10:00–10:20	Coffee Break	
10:20–11:40	Session: Humans in Autonomous Cars (Chair: Markus Enzweiler)	Mohan Trivedi, Klaus Bengler
11:40–12:00	Workshop Closing	
12:00–13:30	Lunch	

Participants

- Michael Aeberhard
BMW AG – München, DE
- Jan Becker
BOSCH Research Center –
Palo Alto, US
- Klaus Bengler
TU München, DE
- Claus Brenner
Leibniz Univ. Hannover, DE
- Wolfram Burgard
Universität Freiburg, DE
- Erik Coelingh
Volvo Car Corporation –
Göteborg, SE
- Michael Darms
Volkswagen AG – Wolfsburg, DE
- Markus Enzweiler
Daimler AG – Böblingen, DE
- Ryan Eustice
University of Michigan – Ann
Arbor, US
- Uwe Franke
Daimler AG – Sindelfingen, DE
- Dariu M. Gavrila
Daimler R&D – Ulm, DE
- Alex Goldberg
HERE – Carlsbad, US
- Ralf G. Herrtwich
Daimler AG – Böblingen, DE
- Ulrich Hofmann
Audi AG – Ingolstadt, DE
- Michael James
Toyota Research Institute North
America – Ann Arbor, US
- Serge Lambermont
Delphi Labs –
Mountain View, US
- Antonio M. Lápez Pena
Autonomus University of
Barcelona, ES
- Chris Mansley
BOSCH Research Center –
Palo Alto, US
- Markus Maurer
TU Braunschweig, DE
- Karsten Mühlmann
Robert Bosch GmbH –
Heilbronn, DE
- Urs Muller
NVIDIA Corp. –
Morganville, US
- Michel Parent
INRIA – Le Chesnay, FR
- Mikael Persson
Linköping University, SE
- Raul Rojas
FU Berlin, DE
- Torsten Sattler
ETH Zürich, CH
- Steven E. Shladover
Univ. of California, Berkeley, US
- Christoph Stiller
KIT – Karlsruher Institut für
Technologie, DE
- Matthias Strauß
Continental Teves AG –
Frankfurt, DE
- Mohan Trivedi
University of California, San
Diego – La Jolla, US
- Sadayuki Tsugawa
AIST – Ibaraki, JP
- D. Scott Williamson
HERE – Chicago, US
- Hermann Winner
TU Darmstadt, DE
- Hans-Joachim Wünsche
Universität der Bundeswehr –
München, DE

Report of Dagstuhl Seminar 15471

Symbolic Computation and Satisfiability Checking

Edited by

Erika Ábrahám[1], Pascal Fontaine[2], Thomas Sturm[3], and Dongming Wang[4]

1 **RWTH Aachen, DE,** abraham@cs.rwth-aachen.de
2 **LORIA – Nancy, FR,** pascal.fontaine@inria.fr
3 **MPI für Informatik – Saarbrücken, DE,** sturm@mpi-inf.mpg.de
4 **Beihang University – Beijing, CN,** dongming.wang@lip6.fr

Abstract

The seminar focused on satisfiability checking for combinations of first-order logic and subclasses thereof with arithmetic theories in a very liberal sense, also covering quantifiers and parameters. It gathered members of the two communities of symbolic computation (or computer algebra) and satisfiability checking (including satisfiability modulo theories). Up-to-now, these two communities have been working quite independently. We are confident that the seminar will initiate cross-fertilization of both fields and bring improvements for both satisfiability checking and symbolic computation, and for their applications.

Seminar November 15–20, 2015 – http://www.dagstuhl.de/15471
Keywords and phrases algorithmic algebra, arithmetic, automated reasoning, decision procedures, quantifier elimination, satisfiability checking, SMT solving, symbolic computation
Digital Object Identifier 10.4230/DagRep.5.11.71
Edited in cooperation with Tim King

1 Executive Summary

Erika Ábrahám
Pascal Fontaine
Thomas Sturm
Dongming Wang

The seminar focused on satisfiability checking for combinations of first-order logic and subclasses thereof with arithmetic theories in a very liberal sense, also covering quantifiers and parameters.

The development of decision procedures for corresponding theories started in the early 20th century in the area of mathematical logic. In the second half of the 20th century it played a prominent role within the development of algebraic model theory. Finally, around 1970, one important research line, viz. algebraic decision methods for real arithmetics, shifted its focus from theoretical results towards practically feasible procedures. That research line was one of the origins of an area known today as *symbolic computation* or *computer algebra*.

More recently, the *satisfiability checking* community, which originated from propositional SAT solving and which is surprisingly disconnected from symbolic computation, began to develop highly interesting results with a particular focus on existential decision problems, following the track of SAT solving towards industrial applications. Powerful *satisfiability*

Symbolic Computation and Satisfiability Checking, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 71–89
Editors: Erika Ábrahám, Pascal Fontaine, Thomas Sturm, and Dongming Wang
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

modulo theories (SMT) solvers were developed, which enrich propositional SAT solving with components for different theories. We understand satisfiability checking in a broad sense, covering besides SMT solving also *theorem proving* with arithmetic.

The two communities of *symbolic computation* and *satisfiability checking* have been quite disjoint, despite strong reasons for them to discuss together. The communities share interests, e.g., examining arithmetic expressions, that are central to both. As a matter of fact, the symbolic computation community has been mostly unaware of basic insights in the satisfiability checking community, such as the efficiency of conflict-driven search with learning, as well as of their fundamental requirements, e.g., incrementality or explanations in the unsatisfiable case. Vice versa, researchers in satisfiability checking have adopted decision procedures from symbolic computation, such as CAD for real closed field, only quite naively, so that they do not really benefit from the considerable experience gained by the original community during 45 years. It is our hope that our seminar contribute to bringing the two communities together, and that they will be much stronger at tackling problems that currently defeat them both, separately.

The seminar offered its participants an opportunity to exchange knowledge about existing methods and applications, to push forward the communication of needs and interests, and to draw attention to challenging open research questions. The participants included researchers from all relevant research areas and with affiliations in academia and as well as in industry. The program was a balanced combination of presentations and tutorials, but also offering time for small group discussions and exchange of ideas.

To the best of our knowledge, the seminar was the first global meeting of the two communities of symbolic computation and satisfiability checking. We are confident that it will initiate cross-fertilization of both fields and bring improvements for both satisfiability checking and symbolic computation, and for their applications.

2 Table of Contents

Executive Summary

Overview of Talks

Panel discussions

3 Overview of Talks

3.1 Quick Intro to CoCoA/CoCoALib

John Abbott (Universität Kassel, DE)

Joint work of Anna Maria Bigatti
URL http://cocoa.dima.unige.it/

The CoCoA software suite implements many algorithms in the realm of Computational Commutative Algebra (especially ideals in multivariate polynomial rings). These implementations are accessible both via a user-friendly interactive system (called CoCoA-5), and more directly as functions in an open-source (GPL3) C++ library (called CoCoALib). There is also a prototype CoCoAServer which uses an OpenMath-like language.

Great emphasis has been placed on making the C++ library easy to use for mathematicians without requiring that they learn advanced features of the C++ language (though such features are used "invisibly" inside the CoCoALib implementations). CoCoALib comes with extensive documentation including around 100 illustrative example programs. CoCoALib has also been interfaced with a number of other specialized C++ libraries (including Normaliz, Frobby and GFan); the design of CoCoALib deliberately aims to accommodate such "collaborations".

To help understand how simple it is to use CoCoALib, we illustrate how to derive Heron's Formula for the area of a triangle, firstly using the CoCoA-5 interactive system (just 8 lines of program code), then using CoCoALib (just 9 lines of C++ beyond the standard "boilerplate" code). This particular example includes the computation of a Groebner basis.

For those with little experience of symbolic computation, we point out some of the strengths and weaknesses which seem most relevant to SMT computations.

An interesting strength is the ability to factorize polynomials. It is even not too costly to find all irreducible factors over the rationals (e.g. CoCoA takes about 1 millisecond for a degree 10 polynomial), so a strategy of "speculative factorization" might be worth considering. There are algorithms for factorizing over algebraic fields, but these are more costly.

Using symbolic computation we can compute exactly in algebraic extensions: numbers are represented as symbolic expressions such as $12\sqrt{2} - 17$, so with this approach there are no problems of "loss of precision". In contrast, arithmetic with such symbolic expressions is relatively costly (especially if the extension degree is high).

To help relate an algebraic number to the real world, symbolic computation also includes techniques for computing guaranteed approximations to solutions of systems of polynomial equations – the easiest case is just a single polynomial. Currently CoCoA can compute approximations only to real roots; it can handle polynomials of degree up to 100 in just a few seconds. Very close approximations can also be obtained relatively quickly.

One important point to note is that there is often a large difference in computation time between a result guaranteed to be absolutely correct, and a result which is correct with very high probability (e.g. 1-epsilon where epsilon is less than 10^{-15}): a typical example is testing a number for primality. Truly guaranteed results would require certified source code, a certified compiler, and certified hardware; with this viewpoint perhaps a probably correct result is more acceptable?

CoCoA also offers "heuristically guaranteed" floating-point arithmetic (called Twin-Floats). This can allow faster computation with real algebraic numbers, but the correctness of the final result is only "probable" – however greater probability incurs greater computational cost. Note that, if asked, CoCoA can recognize when a twin-float represents a simple rational number; more generally it can also find the simplest rational number in a given real interval.

References

1 J. Abbott, A. Bigatti. CoCoALib: A C++ Library for Computations in Commutative Algebra . . . and Beyond. In *Proceedings ICMS 2010*, pages 73–76, 2010.

2 J. Abbott, A. Bigatti. What Is New in CoCoA? In *Proceedings ICMS 2014*, pages 352–358, 2014.

3 J. Abbott, A. M. Bigatti, C. Soeger. Integration of Libnormaliz in CoCoALib and CoCoA-5. In *Proceedings ICMS 2014*, pages 647–653, 2014.

3.2 Building Bridges between Symbolic Computation and Satisfiability Checking

Erika Ábrahám (RWTH Aachen, DE)

Main reference E. Ábrahám, "Building Bridges between Symbolic Computation and Satisfiability Checking," in Proc. of the 2015 ACM on Int'l Symp. on Symbolic and Algebraic Computation (ISSAC'15), pp. 1–6, ACM, 2015.
URL http://dx.doi.org/10.1145/2755996.2756636

The satisfiability problem is the problem of deciding whether a logical formula is satisfiable. For first-order arithmetic theories, in the early 20th century some novel solutions in form of decision procedures were developed in the area of mathematical logic. With the advent of powerful computer architectures, a new research line started to develop practically feasible implementations of such decision procedures. Since then, symbolic computation has grown to an extremely successful scientific area, supporting all kinds of scientific computing by efficient computer algebra systems.

Independently, around 1960 a new technology called SAT solving started its career. Restricted to propositional logic, SAT solvers showed to be very efficient when employed by formal methods for verification. It did not take long till the power of SAT solving for Boolean problems had been extended to cover also different theories. Nowadays, fast SAT-modulo-theories (SMT) solvers are available also for arithmetic problems.

Due to their different roots, symbolic computation and SMT solving tackle the satisfiability problem differently. In this talk we illustrated SMT solving techniques to introduce them to the Symbolic Computation Community, discussed differences and similarities in their approaches, highlighted potentials of combining their strengths, and showed up the challenges that come with this task.

3.3 Open Non-uniform Cylindrical Algebraic Decomposition

Christopher W. Brown (U.S. Naval Academy – Annapolis, US)

Computation with real polynomial equalities and inequalities is an area that has already seen productive interaction between the SMT and computer algebra communities. Many presentations throughout the week highlighted this. For example, Viorica Sofronie-Stokkermans' presentation described how computer algebra software for such computations were used as part of an SMT solving approach to problems in hybrid and reactive systems. This talk concerns interaction the other way, with ideas from the SMT community leading to new results in computer algebra, specifically new results in computing with real polynomial equalities and inequalities. In another presentation, Dejan Jovanovic described joint work with Leonardo de Moura that adapted Cylindrical Algebraic Decomposition (CAD) in a novel way as part of an SMT-style SAT solver for real polynomial equalities and inequalities. One of the key insights of that work was that in following an SMT-style approach, one incrementally builds a model (in this case a point in real space), and that model can be used to optimize certain aspects of CAD construction. This talk describes Non-uniform Cylindrical Algebraic Decomposition (NuCAD), a new kind of CAD that is inspired by their work. It takes the idea of the "model point" and how it can optimize certain aspects of CAD construction, and uses it to to construct NuCADs. As the talk describes, NuCADs can be constructed particularly efficiently, and they can represent solutions of sets of polynomial equalities and inequalities with far fewer cells than CADs.

3.4 Computer Algebra for SAT People

James H. Davenport (University of Bath, GB)

Joint work of R. J. Bradford, M. England, S. McCallum, D. J. Wilson
Main reference M. England, R. J. Bradford, J. H. Davenport, "Improving the Use of Equational Constraints in Cylindrical Algebraic Decomposition," in Proc. of the 2015 ACM Int'l Symp. on Symbolic and Algebraic Computation (ISSAC'15), pp. 165–172, ACM, 2015.
URL http://dx.doi.org/10.1145/2755996.2756678

A brief introduction to computer algebra, focusing on basic tools and on cylindrical algebraic decomposition.

Complexity of dense polynomial arithmetic is well-understood, sparse (and more realistic) polynomial arithmetic less so [4]. Despite the theoretical advances, factoring is still best avoided, and implementations try to: replacing "irreducible" by "square-free and relatively prime" where possible

Cylindrical algebraic decomposition has its roots in [3] who first defined them and gave the first algorithm. Since then, many improvements on this ([5] for the latest), and alternative algorithms base don Regular Chains [2] or Comprehensive Gröbner Bases [6]. The complexity is doubly-exponential in the number of variables, and [1] shows this is inherent, but also that some examples might have this complexity for one variable order, and trivial for a different order.

References

1 C. W. Brown and J. H. Davenport. The Complexity of Quantifier Elimination and Cylindrical Algebraic Decomposition. In C. W. Brown, editor, *Proceedings ISSAC 2007*, pages 54–60, 2007.

2 C. Chen and M. Moreno Maza. An Incremental Algorithm for Computing Cylindrical Algebraic Decompositions. In R. Feng, W.-S. Lee, and Y. Sato, editors, *Proceedings Computer Mathematics ASCM 2009 and 2012*, pages 199–221, 2014.

3 G. E. Collins. Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition. In *Proceedings 2nd. GI Conference Automata Theory & Formal Languages*, pages 134–183, 1975.

4 J. H. Davenport and J. Carette. The Sparsity Challenges. In S. Watt *et al.*, editor, *Proceedings SYNASC 2009*, pages 3–7, 2010.

5 M. England, R. Bradford, and J. H. Davenport. Improving the Use of Equational Constraints in Cylindrical Algebraic Decomposition. In D. Robertz, editor, *Proceedings ISSAC 2015*, pages 165–172, 2015.

6 R. Fukasaku, H. Iwane, and Y. Sato. Real Quantifier Elimination by Computation of Comprehensive Gröbner Systems. In D. Robertz, editor, *Proceedings ISSAC 2015*, pages 173–180, 2015.

3.5 Introduction to Floating-Point Arithmetic

James H. Davenport (University of Bath, GB)

This was a brief introduction to floating-point arithmetic (restricted to IEEE [4], where numbers are $\pm m2^e$ and, generally, $m \in [1, 2)$ is a 53-bit number and $-1022 \le e \le 1023$, so the largest finite number is $> 10^{308}$) from an algebraic point of view. If $\cdot$ is one of the arithmetic operators $\{+, -, \times, /\}$ and $\odot$ the floating-point equivalent, then $a \odot b$ is defined to be $[a \cdot b]$ where $[\cdots]$ denotes the nearest representable number (with precise rules for tie-breaking etc.). Therefore $\oplus$ and $\otimes$ are commutative, and $a \ominus b = \ominus(b \ominus a)$. Admittedly $a \oslash b \neq 1 \oslash (b \oslash a)$ and $a \otimes (b \otimes c) \neq (a \otimes b) \otimes c$, but the differences are trivial (at least for "sensibly-sized" numbers). What, then, are the problems? This list is not exhaustive, and the first three, at least, would hold for any floating-point system.

1. Computing other functions, e.g. $\odot$ when $\cdot \in \{\sin, \log, \ldots\}$, is much harder, as in general we have no idea how many extra places we need to compute with internally. This is a problem known as the "Table Maker's Dilemma" [3]. If we relax $[\cdots]$ to denotes the nearest representable number or its neighbor, then the problem is much more tractable for computation, but much messier for verification, as well as being non-deterministic.
2. $\oplus$ is wildly non-associative: $(1 \oplus 10^{20}) \oplus (-10^{20}) = 0$ but $1 \oplus (10^{20} \oplus (-10^{20})) = 1$. Slightly more subtly, $(5 \oplus 2^{53}) \oplus (-2^{53}) = 4$.

⋆ *Some* languages have rules about what re-arrangements are permissible, and *some* compilers adhere to these, at least with *some* compiler options [5].

3. "sensibly-sized" is far easier to prescribe than to define precisely, or check for.
4. Operations that we would generally think of as "errors" such as $10^{200} \otimes 10^{200}$, generate either representations of $\pm\infty$ ($e = 1024$, $m = 0$), or, as in $\infty - \infty$, "Not a Number" ($e = 1024$, $m \neq 0$) by default, rather than causing an error.

⋆ The speaker recalled seeing a supermarket where the "unit prices", instead of being, say `69 p/litre` were all `NaN p/litre`.

5. As we get very close to 0, there are concepts of "denormalised numbers" and "gradual underflow", which give results with less precision than usual, rather than just 0, which are generally good for the numerical results but hard to model formally (bit-blasting?).
6. Every number is signed, including zero. While this has theoretical advantages when treating complex functions and branch cuts [2], and means that $1 \oslash (1 \oslash x) = x$ when $x = \pm\infty$, it can complicate other areas: $xy \Rightarrow 1/x = 1 \oslash y$is not true when $x = +0$ and $y = -0$.

See, e.g., [3] in general, and [1] about verification.

References

1 M. Brain, C. Tinelli, P. Rümmer, and T. Wahl. An Automatable Formal Semantics for IEEE-754 Floating-Point Arithmetic. http://smtlib.cs.uiowa.edu/papers/BTRW14.pdf, 2014.

2 J. H. Davenport, R. Bradford, M. England, and D. Wilson. Program Verification in the presence of complex numbers, functions with branch cuts etc. In *Proceedings SYNASC 2012*, pages 83–88, 2012.

3 D. Goldberg. What Every Computer Scientist Should Know About Floating-Point Arithmetic. *ACM Comp. Surveys*, 23:5–48, 1991.

4 IEEE. IEEE Standard for Floating-Point Arithmetic (754-2008). *IEEE*, 2008.

5 Intel (Martyn Corden). Consistency of Floating-Point Results using the Intel® Compiler. http://software.intel.com/en-us/articles/consistency-of-floating-point-results-using-the-intel-compiler/, 2012.

3.6 Speed Maple

Jürgen Gerhard (Maplesoft – Waterloo, CA)

Main reference L. Bernardin, P. Chin, P. DeMarco, K. O. Geddes, D. E. G. Hare, K. M. Heal, G. Labahn, J. P. May, L. McCarron, M. B. Monagan, D. Ohashi, S. M. Vorkoetter, “Maple Programming Guide,” Maplesoft, 2015.
URL http://www.maplesoft.com/documentation_center

A brief introduction to the internals of Maple’s mathematical engine was given. Maple’s architecture, the available data types, and the algorithms implemented for polynomial and semi-algebraic systems, and beyond, were summarized.

3.7 Adapting Real Quantifier Elimination Methods for Conflict Set Computation

Maximilian Jaroschek (MPI für Informatik – Saarbrücken, DE), Pascal Fontaine (LORIA – Nancy, FR), and Pablo Federico Dobal (LORIA – Nancy, FR)

Main reference M. Jaroschek, P. F. Dobal, P. Fontaine, “Adapting Real Quantifier Elimination Methods for Conflict Set Computation,” in Proc. of the 10th Int’l Symp. on Frontiers of Combining Systems (FroCos’15), LNCS, Vol. 9322, pp. 151–166, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-319-24246-0_10

The satisfiability problem in real closed fields is decidable. In the context of satisfiability modulo theories, the problem restricted to conjunctive sets of literals, that is, sets of polynomial constraints, is of particular importance. One of the central problems is the computation of good explanations of the unsatisfiability of such sets, i.e. obtaining a small subset of the input constraints whose conjunction is already unsatisfiable. We adapt two commonly used real quantifier elimination methods, cylindrical algebraic decomposition and virtual substitution, to provide such conflict sets and demonstrate the performance of our method in practice.

3.8 Turning CAD Upside Down

Dejan Jovanovic (SRI – Menlo Park, US)

We present the NLSAT decision procedure for solving the existential fragment of non-linear arithmetic. The classic approach to the problem is to project the problem polynomials, followed by model construction (lifting). The new procedure performs the lifting optimistically, and performs focused model-based projection only on the polynomials that are relevant in inconsistencies. The new approach, and the leaner projection operator, are effective in practice, which we support with an extensive experimental evaluation of the implementations in Yices2 and Z3 SMT solvers.

3.9 Constructing a Single CAD Cell

Marek Kosta (MPI für Informatik – Saarbrücken, DE) and Christopher W. Brown (U.S. Naval Academy – Annapolis, US)

Main reference C. W. Brown, M. Kosta, "Constructing a single cell in cylindrical algebraic decomposition," Journal of Symbolic Computation, Vol. 70, pp. 14–48, 2015.
URL http://dx.doi.org/10.1016/j.jsc.2014.09.024

We present an algorithm which, given a point and a set of polynomials, constructs a single cylindrical cell containing the point, such that the polynomials are sign-invariant in the computed cell. To represent a single cylindrical cell, a novel recursive data structure is introduced. The algorithm works with the data structure and proceeds by incrementally merging the input polynomials into the cell. A merge procedure realizing this refinement is described. The merge procedure is based on McCallum's operator, but uses geometric information relative to the test point to reduce the projection set. The use of McCallum's operator implies the incompleteness of our algorithm in general. However, the algorithm is complete for well-oriented sets of polynomials. Moreover, the incremental approach described can be easily adapted to a different projection operator. Our cell construction is an alternative to the "model-based" method described by D. Jovanovié during this seminar.

3.10 Symbol Elimination for Program Analysis

Laura Kovács (Chalmers UT – Göteborg, SE)

Joint work of Laura Kovács and Andrei Voronkov
Main reference L. Kovács, A. Voronkov, "Finding Loop Invariants for Programs over Arrays Using a Theorem Prover," in Proc. of the 12th Int'l Conf. on Fundamental Approaches to Software Engineering (FASE'09), LNCS, Vol. 5503, pp. 470–485, Springer, 2009.
URL http://dx.doi.org/10.1007/978-3-642-00593-0_33

I describe our new symbol elimination method [1, 2] for generating and proving properties about software systems. Symbol elimination uses first-order theorem proving and symbolic computation techniques to automatically discover non-trivial program properties, such as loop invariants and loop bounds. Moreover, symbol elimination can be used as an alternative to interpolation for software verification. The talk will describe how symbol elimination can be used for polynomial and quantified invariant generation, by using Groebner basis computation, quantifier elimination and saturation-based first-order theorem proving. Symbol elimination is implemented in the award-winning first-order theorem prover Vampire and successfully evaluated on a large number of examples coming from academic and industrial benchmarks.

References

1 Laura Kovács and Andrei Voronkov. *Finding Loop Invariants for Programs over Arrays Using a Theorem Prover*. FASE 2009: 470-485.

2 Laura Kovács and Andrei Voronkov. *Interpolation and Symbol Elimination*. CADE 2009: 199-213.

3.11 SMT-RAT: An Open Source C++ Toolbox for Strategic and Parallel SMT Solving

Gereon Kremer (RWTH Aachen, DE) and Florian Corzilius (RWTH Aachen, DE)

Main reference F. Corzilius, G. Kremer, S. Junges, S. Schupp, E. Ábrahám, "SMT-RAT: An Open Source C++ Toolbox for Strategic and Parallel SMT Solving," in Proc. of the 18th Int'l Conf. on Theory and Applications of Satisfiability Testing (SAT'15), LNCS, Vol. 9340, pp. 360–368, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-319-24318-4_26

During the last decade, popular SMT solvers have been extended step-by-step with a wide range of decision procedures for different theories. Some SMT solvers also support the user-defined tuning and combination of such procedures, typically via command-line options. However, configuring solvers this way is a tedious task with restricted options.

In this paper we present our modular and extensible C++ library SMT-RAT, which offers numerous parameterized procedure modules for different logics. These modules can be configured and combined into an SMT solver using a comprehensible whilst powerful strategy, which can be specified via a graphical user interface. This makes it easier to construct a solver which is tuned for a specific set of problem instances. Compared to a previous version, we have extended our library with a number of new modules and support for parallelization in strategies. An additional contribution is our thread-safe and generic C++ library CArL, offering efficient data structures and basic operations for real arithmetic, which can be used for the fast implementation of new theory-solving procedures.

3.12 An Invitation to Numerical Algebraic Geometry

Viktor Levandovskyy (RWTH Aachen, DE)

Joint work of Levandovskyy, Viktor; Hauenstein, Jonathan
Main reference J. D. Hauenstein, V. Levandovskyy, "Certifying solutions to square systems of polynomial-exponential equations," arXiv:1109.4547v1 [math.NA], 2011; to appear in Journal of Symbolic Computation.
URL http://arxiv.org/abs/1109.4547v1

In this ad-hoc contributed talk we present some basic facts from the emerging theory called "Numerical Algebraic Geometry". In particular, the results from the Smale's alpha-theory will be explained. There is an algorithm, allowing one to certify a solution of a system of nonlinear polynomials equations. Moreover, for some number of close candidate solutions, it is possible either to distinguish their associated solutions or to conclude that there is one associated solution with multiplicity greater than one. Since very recently there evolve more and more advanced algorithms, based on these two fundamental ones. These algorithms are often implemented in a freely available package "alphaCertified" and use a freely available system "bertini". A live demonstration of the former is presented as well.

3.13 A Presentation of Satisfiability Modulo Theory for Nonspecialists

David Monniaux (VERIMAG – Grenoble, FR)

Satisfiability modulo theory (SMT) extends propositional satisfiability (SAT) by allowing arithmetic predicates, or more generally predicates within a theory. $a \wedge (b \vee \neg c)$ is a SAT formula whose solutions are $(a = \text{true}, b = \text{true}, c = \text{true})$, $(a = \text{true}, b = \text{true}, c = \text{false})$, $(a = \text{true}, b = \text{false}, c = \text{false})$. $(x > 0) \wedge (y > 0) \wedge (x + y < 1)$ is an SMT formula over linear real arithmetic (LRA) whose solutions include e.g. $(x = \frac{1}{4}, y = \frac{1}{4})$, but the same formula admits no solution over linear integer arithmetic (LIA). $f(x) \neq f(y) \wedge x = y$, where f stands for an unspecified ("uninterpreted") function (UF) from integers to integers, has no solution since if $x = y$, for any function f, $f(x) = f(y)$. Examples of SMT theories include LRA, LIA, UF, and combinations thereof. SMT solvers may allow quantifiers inside formulas, though these quickly make the problem undecidable (e.g. UF+LIA+$\forall$ is undecidable).

The most common way of implementing SMT is by the "DPLL(T)" framework, in which a SAT solver based on the CDCL (constraint-driven clause learning) principle (a modern evolution of Davis-Putnam-Logemann-Loveland) interacts with a decision procedure for conjunctions of predicates from the theory. The formula is stripped to its propositional structure, e.g. $(x > 0) \wedge (y > 0) \wedge (x + y < 1 \vee y < 0)$ is replaced by $a \wedge b \wedge (c \vee d)$ where a stands for $x > 0$, b for $y > 0$, c for $x + y < 1$ and d for $y < 0$. An assignment $a = \text{true}$, $b = \text{true}$, $c = \text{true}$ is made but the decision procedure for LIA informs the SAT solver than this assignment is inconsistent. A clever decision procedure will strive to inform the SAT solver than a sub-part of the assignment only is inconsistent, since this will rule out a larger part of the propositional state space.

More recently, alternatives to DPLL(T) such as MCSAT or the extension of DPLL to values outside of Booleans have been proposed. They have in common that they reason directly about rational, integer etc. values instead of going through a propositional abstraction.

Satisfiability testing is typically used in program analysis to prove that a finite sequence (including tests and other control flow) of statements is infeasible (in program verification), or to find concrete values going through that sequence (in automated test case generation). The use cases can be more complicated; for instance satisfying assignments may be used to automatically refine the search for inductive invariants in program verification.

Finally, Craig interpolation is another outcome of satisfiability testing. Given $A(x, y)$, $B(y, z)$ and $C(z, t)$, a Craig interpolant is a pair (I, J) such that for all x, y, z, t:

$$A(x, y) \implies I(y), I(y) \wedge B(y, z) \implies J(z), J(z) \wedge C(z, t) \implies \text{false}.$$

Such interpolants are used in program analysis to give "local arguments" why a trace with steps A,B,C is infeasible. Finding simple interpolants likely to generalize to other traces and becoming inductive is an active area of research.

3.14 Triangular Sets over F2 vs. Satisfiability Checking: A Potential Connection and Interaction?

Chenqi Mou (Beihang University – Beijing, CN) and Dongming Wang (Beihang University – Beijing, CN)

In this talk the concepts of triangular sets and triangular decomposition are presented. In particular, a typical splitting strategy in triangular decomposition and then a refined and easier one in Boolean rings are illustrated, with efforts to reveal the similarity between Boolean triangular sets and SAT solving.

References

1 X.-S. Gao and Z. Huang. Characteristic set algorithms for equation solving in finite fields In *Journal of Symbolic Computation, 47(6)*, pages 655–679, 2012.

2 M. Kalkbrener. A generalized Euclidean algorithm for computing triangular representations of algebraic varieties In *Journal of Symbolic Computation, 15(2)*, pages 143–167, 1993.

3 D. Wang. Decomposing polynomial systems into simple systems In *Journal of Symbolic Computation, 25(3)*, pages 295–314, 2000.

4 D. Wang. Computing triangular systems and regular systems In *Journal of Symbolic Computation, 30(2)*, pages 221–236, 2000.

5 W.-T. Wu. On zeros of algebraic equations: An application of Ritt principle In *Kexue Tongbao, 31(1)*, pages 1–5, 1986.

6 L. Yang and J.-Z. Zhang. Searching dependency between algebraic equations: An algorithm applied to automated reasoning In *Artificial Intelligence in Mathematics*, pages 147–156, 1994.

3.15 raSAT: SMT Solver for Nonlinear Arithmetic

Mizuhito Ogawa (JAIST – Ishikawa, JP)

We present the raSAT SMT solver for polynomial constraints, which aims to handle them over both reals and integers with simple unified methodologies: (1) raSAT loop for inequalities, which extends the interval constraint propagation with testing to accelerate SAT detection, and (2) a non-constructive reasoning for equations over reals, based on the generalized intermediate value theorem. raSAT has participated SMT-COMP 2015, and was 3rd (among 6) in QF_NRA and 2nd (among 7) in QF_NIA categories of main tracks.

3.16 Using Instantiation-Based Approaches for Quantifier Elimination in SMT

Andrew Joseph Reynolds (EPFL – Lausanne, CH)

This talk presents instantiation-based decision methods for determining the satisfiability of quantified formulas in first-order theories. Using this framework, we obtain decision procedures for linear real arithmetic (LRA) and linear integer arithmetic (LIA) formulas with one quantifier alternation. Our procedure can be integrated into the solving architecture used by typical SMT solvers. Experimental results on standardized benchmarks from model checking, static analysis, and synthesis show that our implementation of the procedure in the SMT solver CVC4 outperforms existing approaches for quantified linear arithmetic.

3.17 How Far We Can Eliminate Quantified Variables Using Equalities

Yosuke Sato (Tokyo University of Science, JP)

When a given quantified formula contains equalities not only explicitly but also implicitly, we can use them for eliminating quantifies. In order to use all such equalities we need computation of comprehensive Groebner bases. In the talk we introduce our work which gives a sufficiently practical such method.

3.18 Introduction to iSAT3

Karsten Scheibler (Universität Freiburg, DE)

Joint work of Herde, Christian; Kupferschmid, Stefan; Teige, Tino; Eggers, Andreas; Neubauer, Felix; Mahdi, Ahmed

We present iSAT3, a satisfiability checker for Boolean combinations of arithmetic constraints over real- and integer-valued variables. The solver supports constraints containing linear and non-linear arithmetic as well as transcendental functions. iSAT3 tightly integrates

Interval-Constraint-Propagation (ICP) into the Conflict-Driven Clause-Learning (CDCL) framework.

3.19 Hierarchical Reasoning for the Verification of Parametric Systems

Viorica Sofronie-Stokkermans (Universität Koblenz-Landau, DE)

We show how hierarchical reasoning and quantifier elimination can be used to automatically provide guarantees that given parametric systems satisfy certain safety or invariance conditions. Such guarantees can be expressed as constraints on parameters.

We present several examples (from the verification of reactive (or discrete time) systems and of linear hybrid automata) and point out some challenges we encountered when using existing systems for quantifier elimination.

3.20 The Rodin Platform and SMT Solvers

Laurent Voisin (SYSTEREL Aix-en-Provence, FR)

Event-B is a formal notation for modelling discrete transition systems. It is based on first-order predicate calculus with equality completed with typed set theory and integer arithmetics. The system is modelled by a state and its properties as invariants. Transitions are described by guarded events that can fire atomically as soon as their guard holds.

The Rodin platform is the reference tool for modelling with the Event-B notation and proving the model correct. It has been developed for more than ten years and is freely available from http://event-b.org under an Eclipse license. The main design principles of the platform are openness (open source and open syntax), extensibility (more than thirty plug-ins are available) and a reactive modelling process (tools are run automatically in the background).

The interactive prover of the Rodin platform is tactic based. The prover just maintains a proof tree in the sequent calculus. The rules of the proof tree are not predefined, but produced by reasoners (which are invoked by the tactics). The reasoners can be either internal (provided by the core platform) or external (provided by plug-ins). In order to foster reuse after model modification (which always happens), the proof rules shall be as small as possible.

The SMT plug-in works as follows: It first translates an Event-B sequent to its negation in the SMT-LIB format, then invokes an SMT solver. If the solver returns SAT, then the sequent is not valid and the reasoner fails. If the solver returns UNSAT, then the sequent is valid and the UNSAT core provided by the solver is used to build the proof rule.

Two examples are particularly interesting. One is a simple property of a barycenter. It consists in proving that

$$\begin{array}{ll} & 0 \leq N \ \wedge \ 0 \leq a \ \wedge \ 0 \leq b \\ & x \in 0\,..\,N \ \wedge \ y \in 0\,..\,N \\ \vdash & (a * x + b * y) \div (a + b) \in 0\,..\,N \end{array}$$

No solver currently connected to the Rodin platform can prove this property, as it uses non-linear arithmetic. Consequently, one has to prove it manually by unfolding the definitions of the operators, which is quite tedious.

A second example consists in stating that an acylic relation is non-reflexive:

$$\begin{array}{ll} & r \in S \leftrightarrow S \\ & \forall p \cdot p \subseteq r^{-1}[p] \implies p = \emptyset \\ \vdash & \mathrm{id} \cap r = \emptyset \end{array}$$

Surprisingly, this example is proved by the CVC3 solver, although it needs second order instantiation to find the right instance for p.

3.21 The SMT Theory of Floating-Point Arithmetic

Christoph M. Wintersteiger (Microsoft Research UK – Cambridge, GB)

URL http://smtlib.cs.uiowa.edu/theories-FloatingPoint.shtml

The Satisfiability Modulo Theories (SMT) community recently added a standard for floating-point arithmetic (SMT FP) to its set of theories. This enables many verification techniques that already build on SMT solvers to add support for floating-point arithmetic at a very low cost. In this presentation I summarize and demonstrate the scope of the new theory; it is, to a large extent, based on the IEEE-754 standard, but there are some intricacies in which SMT FP departs from IEEE-754 in an effort to simplify and streamline common problems, and I will point some of those intricacies out. Support for SMT FP in actual SMT solvers exists, but is not commonplace yet; those solvers that support it are either pure translation to Boolean logic (bit-blasters), or they employ abstraction refinement schemes that promise to be more efficient on many problems. At the time, the solvers with support for SMT FP are MathSAT, Sonolar, CVC4, and Z3.

3.22 Stability of Parametric Decomposition

Kazuhiro Yokoyama (Rikkyo University – Tokyo, JP)

Main reference Kazuhiro Yokoyama, "Stability of Parametric Decomposition," in Proc. of the 2nd Int'l Congress on Mathematical Software (ICMS'06), LNCS, Vol. 4151, pp. 391–402, Springer, 2006.
URL http://dx.doi.org/10.1007/11832225_39

We deal with ideals generated by polynomials with parametric coefficients, and introduce "stabilities on ideal structures" based on stability of forms of Gröbner bases. Then, we extend those stabilities to radicals and irreducible decompositions and show the computational tractability on those computations by integrating existing techniques.

Making these computational realizations efficient and practical is still ongoing, but it is very challenging for Computer Algebra not only in theory, but also in application. It will certainly help to develop the abilities of Computer Algebra and widen its application, including Quantifier Elimination and Satisfiability Checking.

4 Panel discussions

4.1 SMT Solvers and Computer Algebra Systems: Potentials of Technology Transfer

Erika Ábrahám (RWTH Aachen, DE)

During and after the talk "Building Bridges between Symbolic Computation and Satisfiability Checking", we discussed different issues of connecting algorithms and implementations rooted in the two communities of Symbolic Computation and Satisfiability Checking.

On the one hand, SMT solving has its strength in efficient techniques for exploring Boolean structures, learning, combining solving techniques, and developing dedicated heuristics, but its current focus lies on easier theories and it makes use of symbolic computation results only in a rather naive way. There are fast SMT solvers available for the satisfiability check of linear real and integer arithmetic problems, but just a few can handle non-linear arithmetic.

On the other hand, Symbolic Computation is strong in providing powerful procedures for sets (conjunctions) of arithmetic constraints, but it does not exploit the achievements in SMT solving for efficiently handling logical fragments, using heuristics and learning to speed-up the search for satisfying solutions.

The SMT-solving community could definitely profit from further exploiting Symbolic Computation achievements and adapt and extend them to comply with the requirements on embedding in the SMT context. However, it is a highly challenging task, as it requires a deep understanding of complex mathematical problems, whose embedding in SMT solving is far from trivial and needs their adaptation and extension. Symmetrically, Symbolic Computation could profit from exploiting successful SMT ideas, but it requires expertise in efficient solver technologies and their implementation, like dedicated data structures, sophisticated heuristics, effective learning techniques, and approaches for incrementality and explanation generation in theory solving modules.

We discussed how we could overcome these problems: on the one hand, how to adapt implementations of decision procedures in computer algebra systems to satisfy the requirements for SMT-embedding, supporting incrementality, the generation of models for satisfiable problems and explanations for unsatisfiable problem instances, and to offer suitable interfaces for SMT calls; on the other hand, how to adapt successful SAT and SMT technologies to improve the efficiency of computer algebra systems, making use of efficient handling of logical structures, learning, and developing dedicated search heuristics.

4.2 SMT-LIB: An Introduction

Pascal Fontaine (LORIA – Nancy, FR)

Joint work of Barrett, Clark; Ranise, Silvio; Stump, Aaron; Tinelli, Cesare
Main reference C. Barrett, P. Fontaine, C. Tinelli, "The SMT-LIB Standard: Version 2.5," 2015.
URL http://smtlib.cs.uiowa.edu/language.shtml
URL http://www.SMT-LIB.org

The Satisfiability Modulo Theories (see [2] for a survey on SMT) community is built around the SMT-LIB initiative (http://smtlib.cs.uiowa.edu/). The aim of this initiative was at first to collect a library of benchmarks. The emergence of the SMT-LIB language quickly followed, as a necessity for exchanging benchmarks and unambiguously interpreting them. This also involved being able to describe precisely the underlying concepts behind SMT, i.e. the theories and their combinations. The SMT-LIB standard is constantly improving, and aims at tackling always richer languages while at the same time keeping the standard simple. Nowadays, SMT-LIB is supported by all the main SMT solvers. It is used as the interface language of many tools (e.g. verification platforms) with their SMT solver backends. It is also the official language of the SMT-COMP (http://www.smtcomp.org/), the annual competition of solvers.

We discussed the SMT-LIB language version 2.5 [1], with a focus on arithmetic theories and logics. The discussion was also the opportunity to compare its features with other languages in use in the Symbolic Computation community. The SMT-LIB language version 2.5 is not extendable but version 3.0 should be more flexible, and could accommodate some of the needs for a language suitable in a larger context.

References

1 Clark Barrett, Pascal Fontaine, and Cesare Tinelli. The SMT-LIB Standard: Version 2.5, 2015. Available at http://www.SMT-LIB.org.

2 Clark Barrett, Roberto Sebastiani, Sanjit A. Seshia, and Cesare Tinelli. Satisfiability modulo theories. In Armin Biere, Marijn J. H. Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*, chapter 26, pages 825–885. IOS Press, February 2009.

Participants

- John Abbott
  Universität Kassel, DE
- Erika Ábrahám
  RWTH Aachen, DE
- Bernd Becker
  Universität Freiburg, DE
- Martin Bromberger
  MPI für Informatik –
  Saarbrücken, DE
- Christopher W. Brown
  U.S. Naval Academy –
  Annapolis, US
- Shaowei Cai
  Chinese Academy of Sciences –
  Beijing, CN
- Florian Corzilius
  RWTH Aachen, DE
- James H. Davenport
  University of Bath, GB
- Pascal Fontaine
  LORIA – Nancy, FR
- Stephen Forrest
  Maplesoft Europe GmbH, DE
- Jürgen Gerhard
  Maplesoft – Waterloo, CA
- Maximilian Jaroschek
  MPI für Informatik –
  Saarbrücken, DE
- Dejan Jovanovic
  SRI – Menlo Park, US
- Tim A. King
  Google Inc. –
  Mountain View, US
- Konstantin Korovin
  University of Manchester, GB
- Marek Kosta
  MPI für Informatik –
  Saarbrücken, DE
- Laura Kovács
  Chalmers UT – Göteborg, SE
- Gereon Kremer
  RWTH Aachen, DE
- Wolfgang Küchlin
  Universität Tübingen, DE
- Viktor Levandovskyy
  RWTH Aachen, DE
- Klaus Meer
  BTU Cottbus, DE
- David Monniaux
  VERIMAG – Grenoble, FR
- Chenqi Mou
  Beihang University – Beijing, CN
- Mizuhito Ogawa
  JAIST – Ishikawa, JP
- Andrew Joseph Reynolds
  EPFL – Lausanne, CH
- Yosuke Sato
  Tokyo University of Science, JP
- Karsten Scheibler
  Universität Freiburg, DE
- Tobias Schubert
  Universität Freiburg, DE
- Viorica Sofronie-Stokkermans
  Universität Koblenz-Landau, DE
- Thomas Sturm
  MPI für Informatik –
  Saarbrücken, DE
- Laurent Voisin
  SYSTEREL Aix-en-Provence,
  FR
- Christoph M. Wintersteiger
  Microsoft Research UK –
  Cambridge, GB
- Patrick Wischnewski
  Logic4Business –
  Saarbrücken, DE
- Kazuhiro Yokoyama
  Rikkyo University – Tokyo, JP

Report of Dagstuhl Seminar 15472

Programming with “Big Code”

Edited by

William W. Cohen[1], Charles Sutton[2], and Martin T. Vechev[3]

1 **Carnegie Mellon University, US,** wcohen@cs.cmu.edu
2 **University of Edinburgh, GB,** csutton@inf.ed.ac.uk
3 **ETH Zürich, CH,** martin.vechev@inf.ethz.ch

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15472 *Programming with “Big Code”*. “Big Code” is a term used to refer to the increasing availability of the millions of programs found in open source repositories such as GitHub, BitBucket, and others.

With this availability, an opportunity appears in developing new kinds of statistical programming tools that learn and leverage the effort that went into building, debugging and testing the programs in “Big Code” in order to solve various important and interesting programming challenges.

Developing such statistical tools however requires deep expertise across multiple areas of computer science including machine learning, natural language processing, programming languages and software engineering. Because of its highly inter-disciplinary nature, the seminar involved top experts from these fields who have worked on or are interested in the area.

The seminar was successful in familiarizing the participants with recent developments in the area, bringing new understanding to different communities and outlining future research directions.

Seminar November 15–18, 2015 – http://www.dagstuhl.de/15472

1998 ACM Subject Classification D.2.3 [Software Engineering] Coding Tools and Techniques, I.2.2 [Artificial Intelligence] Automatic Programming, I.2.5 [Artificial Intelligence] Programming Languages and Software

Keywords and phrases machine learning, natural language processing, programming languages, software engineering, statistical programming tools

Digital Object Identifier 10.4230/DagRep.5.11.90

1 Executive Summary

Martin T. Vechev
William W. Cohen
Charles Sutton

The main objective of the seminar was to bring together several research communities which have so far been working separately on the emerging topic of “Big Code” and to foster a new community around the topic. Over the last 4–5 years there have been several developments and interesting results involving “Big Code” all spanning a wide range of fields and conferences: the seminar brought these communities together and enabled them to interact for the first time.

The program was structured as a series of talks interspersed with discussion. Almost all of seminar participants gave a talk on their latest research. Even though the initial plan was to include special discussion sessions, each talk triggered so much discussion, both during the talk itself, and also after, that there was no need for specific discussion slots. We believe the seminar was successful in setting the right atmosphere for open ended discussion and obtained the desired affect of triggering much organic interaction.

Only the last day (morning) included a short wrap-up discussion session focusing on the future of the area, defining common data sets and future challenges the community can address. That discussion is summarized in the working group report.

The seminar was highly inter-disciplinary involving experts from programming languages, software engineering, machine learning and natural language processing. Further, it brought together research groups from Europe, Asia and U.S., all working on the topic of "Big Code", and raised awareness and familiarity with what different research groups are working on.

The talks and discussions spanned several topics including: the kinds of statistical methods used (e.g., n-gram models, recurrent neural networks, graphical models, probabilistic grammars, etc), new programming applications that can benefit from these models (e.g., code completion, code search, code similarity, translating natural language to code, etc), and the interaction between these. Some of the presentations were more of an introductory/overview nature while others focused on the more technical aspects of particular programming tools and machine learning models.

After two days of presentations and discussions, we used the last day of the seminar (before lunch) to summarize the discussions and to outline a future research direction. A suggestion enthusiastically embraced by everyone was to create a web site which lists the current data sets, challenges, tools and research groups working on the topic. The view was that this will not only enable existing groups to compare their tools on common problems and data sets but will also make it much easier for other research groups and graduate students to get into the area and to start contributing. It also serves as a useful instrument for raising awareness about the topic:

We have now created this web site and have made it available here: http://learnbigcode.github.io/.

In a short time, several groups have started contributing by uploading links to tools, data sets and challenges.

Overall, the seminar was successful both in terms of stimulating new and fruitful interaction between research communities that were working in the area but were separated so far, but also in setting a common agenda moving forward. Due to the high interest and feedback from this seminar, we anticipate that in a year or two from now, we will be ready to propose a larger seminar on the topic.

2 Table of Contents

Executive Summary

Overview of Talks

Working groups

3 Overview of Talks

3.1 Miltos Allamanis

Miltos Allamanis (University of Edinburgh, GB) and Charles Sutton (University of Edinburgh, GB)

We briefly discuss recent work on mining code idioms from codebases. A code idiom is a syntactic code fragment that recurs frequently across software projects and has a single semantic purpose. Although, we know that developers write idiomatic code it is not clear why idioms arise in source code. In this talk, I discuss potential reasons for the prevalence of idiomatic code among developers.

3.2 Loop-Invariant Synthesis using Techniques from Constraint Programming

Jason Breck (University of Wisconsin – Madison, US)

In this talk, I describe a loop invariant synthesis technique inspired by constraint programming. In particular, the technique uses an abstract domain from constraint programming: the abstract domain consists of sets of boxes, where a box is a collection of interval constraints, one for each program variable. The technique synthesizes inductive loop invariants for programs that manipulate real-valued variables. It works by iteratively splitting and deleting boxes until the set of boxes becomes an inductive loop invariant, or a failure condition is reached. I describe an extension to the technique that uses an abstract domain of octagons instead of boxes. I also describe a series of experiments that test our technique on programs taken from the literature on numeric loop invariant synthesis.

3.3 An overview of the Pliny project

Swarat Chaudhuri (Rice University – Houston, US) and Christopher M. Jermaine (Rice University – Houston, US)

Formal methods is the science of mechanized formal reasoning about complex systems. Data mining is the science of extracting knowledge and insights from large volumes of data. In this talk, I will describe Pliny, a Rice-led DARPA project that seeks to bridge these two disciplines. The vision of Pliny is to develop a wide range of formal reasoning tools that aim to make software more reliable, faster, and more easily programmed. The difference between Pliny and existing formal methods approaches is that Pliny complements automated logic-based analysis with statistical mining of “Big Code”, i.e., large corpora of open-source software. The logical and statistical techniques are unified under a Bayesian framework where logical techniques are guided by data-driven insights and data mining happens on artifacts generated through automated reasoning.

3.4 Studying the Naturalness of Software

Premkumar T. Devanbu (University of California – Davis, US)

Natural languages have evolved to serve immediate, natural human purposes: survival, nourishment, reproduction; while languages per se are rich in vocabulary and grammatical flexibility, most human utterances are simple and repetitive, reflecting the origins of the medium. At UC Davis, we discovered in 2011 that large software corpora show even greater repetitive nature, despite the considerable power and flexibility of programming languages. Since studies since then, we have studied this repetitive structure in detail, and shown that the phenomenon persists (and indeed strengthens) even if differences between programming language corpora and natural language corpora are accounted for. Thus, although programming languages have much greater vocabulary (arising from variable names) the vocabulary, when names are split, show an even greater degree of repetition. Likewise, the simplicity of programming code is not just an artifact of simpler structure: when compared on an equal basis (programming corpora without keywords and operators, language corpora without function words) software in fact becomes even more repetitive. IN ongoing work, we are finding that when software code is non-repetitive (or surprising), it is in fact much likelier to be defective.

3.5 Doing Software Analytics Research – Incorporating Cross-Domain Expertise

Shi Han (Microsoft Research – Beijing, CN)

This talk introduces four selected research projects in the past six years at the Software Analytics group of Microsoft Research, demonstrating te importance and challenges of incorporating cross-domain expertise for successful learning/mining tasks against software artifacts such as code or log.

- StackMine – performance debugging in the large via mining millions of stack traces
- DriverMine – comprehending OS performance issues in the software ecosystem scope
- JSweeter – uncovering JavaScript performance code smells relevant to type mutation
- Codeology – program understanding via mining big code

3.6 Can big code find errors?

Abram Hindle (University of Alberta – Edmonton, CA)

Syntax errors, misspellings, and misunderstandings are detriment to programmers everywhere. But naturalness is here help, by treating software source code as natural language utterances we can leverage tools used in the NLP domain to help debug software. Specifically in this work we demonstrate that smoothed n-gram models of source code tokens trained on a large corpus of code can easily determine if code doesn’t belong: code that doesn’t belong, code that is surprising to a model is usually code that contains syntax errors, misspellings, or awkward semantics not usually employed. We evaluate this conjecture on code from 2 fundamentally different programming languages: Java and Python. We find that in both cases asking a model trained on good source code, “does this potentially bad source code surprise you?” allows us to identify syntax errors, misnamed identifiers, and even missing code tokens. We find that in the case of Python, a dynamic language, that the lack of oracle is a huge impediment, such that programmers may ship python code that seems to work but actually contains syntax errors.

3.7 Mining and Understanding Software Enclaves

Suresh Jagannathan (Purdue University – West Lafayette, US)

The modern-day software ecosystem is a messy and chaotic one. Among other things, it includes an intricate stack of sophisticated services and components, susceptible to frequent (and often incompatible) upgrades and patches; emerging applications that operate over large, unstructured, and noisy data; and, an ever-growing code base replete with latent defects and redundancies. Devising novel techniques to tame this complexity, and improve software resilience, trustworthiness, and expressivity in the process, is a common theme actively being explored by several ongoing DARPA programs. This talk gives an overview of one such effort – MUSE (Mining and Understanding Software Enclaves) –, which aims to exploit predictive analytics over large software corpora to automatically repair and synthesize programs. It seeks to realize this vision of “Big Code” by developing foundational advances in programming language design, analysis, and implementation, and has as its overarching goal, revolutionizing the way we think about software construction and reliability.

3.8 Learning to Search Large Code Bases

Christopher M. Jermaine (Rice University – Houston, US)

In the Pliny project, our goal is to use a large code repository to help perform tasks such as synthesis, bug finding, and repair. One of the key subtasks for each of these tasks is searching a large database for all codes that are “close to” a query specification. Unfortunately, “close to” is not defined a priori. We can extract features from code (including syntactic and semantic information about the code) and we can define distance metrics over those features, but we do not know beforehand which are going to be most useful for solving a given task. In this talk, I will describe how we plan to use the question-answer posts from Stack Overflow to help bootstrap the process of figuring out what features and metrics are potentially useful to power search. We describe a model that is similar to Canonical Correlation Analysis, and automatically chooses features and metrics that are useful for linking code to natural language text. Our hypothesis is that those same features and metrics will be useful for finding codes that are close to a query in a synthesis, debugging, or repair task.

3.9 Big Code for Better Code

Sebastian Proksch (TU Darmstadt, DE)

The goal of this talk was to stimulate discussion. Because of a very restricted time slot for the presentation, the talk was very focussed on two points:

1. It introduced an extensible inference engine for recommender systems in software engineering that is based on feature vectors generated from structural context. Researchers can freely exchange the underlying pattern detection mechanisms and get a recommender and the necessary evaluation pipeline for free.
2. State of the art evaluation of recommender systems in software engineering are typically based on artificial evaluations. The talk stresses the fact that this does not provide any insights about the perceived usefulness of a tool by the developer. To improve evaluations, we collected real developer feedback by instrumenting the Visual Studio IDE and use this as a ground truth for evaluation instead.

3.10 Language to Code: Learning Semantic Parsers for If-This-Then-That Recipes

Christopher Quirk (Microsoft Corporation – Redmond, US)

Using natural language to write programs is a touchstone problem for computational linguistics. We present an approach that learns to map natural-language descriptions of simple “if-then” rules to executable code. By training and testing on a large corpus of naturally-occurring programs (called “recipes”) and their natural language descriptions, we demonstrate the ability to effectively map language to code. We compare a number of semantic parsing approaches on the highly noisy training data collected from ordinary users, and find that loosely synchronous systems perform best.

3.11 Learning Programs from Noisy Data

Veselin Raychev (ETH Zürich, CH) and Martin T. Vechev (ETH Zürich, CH)

We present a novel technique for constructing statistical code completion systems. These are systems trained on massive datasets of open source programs, also known as “Big Code”. The key idea is to introduce a domain specific language (DSL) over trees and to learn functions in that DSL directly from the dataset. These learned functions then condition the predictions made by the system. This is a flexible and powerful technique which generalizes several existing works as we no longer need to decide a priori on what the prediction should be conditioned. As a result, our code completion system surpasses the prediction capabilities of existing, hard-wired systems.

3.12 Clio:Digital Code Assistant for the Big Code Era

Armando Solar-Lezama (MIT – Cambridge, US)

The talk describes our efforts in leveraging information from big code in order to support synthesis tasks. The first part of the talk first provides a brief overview of our DemoMatch effort which allows user to demonstrate the use of a framework on an existing application and then generates the code necessary to use the framework in that way. The second part of the talk describes a tool called Swapper that automatically generates formula simplification routines to be used inside solvers. Swapper takes as input a corpus of formulas from synthesis problems and produces simplifiers tailored to those formulas. Automatic generation of such simplifiers is a first step towards automatic generation of domain specialized solvers.

3.13 Statistical Analysis of Program Text

Charles Sutton (University of Edinburgh, GB)

Billions of lines of source code have been written, many of which are freely available on the Internet. This code contains a wealth of implicit knowledge about how to write software that is easy to read, avoids common bugs, and uses popular libraries effectively.

We want to extract this implicit knowledge by analyzing source code text. To do this, we employ the same tools from machine learning and natural language processing that have been applied successfully to natural language text. After all, source code is also a means of human communication.

We present three new software engineering tools inspired by this insight, that learn local coding conventions (naming and formatting), syntactic idioms in code, and API patterns.

3.14 Graph-structured Neural Networks for Program Verification

Daniel Tarlow (Microsoft Research UK – Cambridge, GB)

An open problem in program verification is to verify properties of computer programs that manipulate memory on the heap. A key challenge is to find loop invariants – formal descriptions of the data structures that are instantiated – which are used as input to a proof procedure that verifies the program. We describe a machine learning-based approach, where we execute the program and then learn to map the state of heap memory (represented as a labelled directed graph) to a logical description of the instantiated data structures. In the process of working on this problem, we developed a new general purpose neural network architecture that is suitable for learning mappings from graph-structured inputs and sequential outputs. We describe this model and speculate that it could be generally useful for a range of problems that arise in the space of “big code”.

3.15 Learning from Big Code

Martin T. Vechev (ETH Zürich, CH)

An overview presentation covering the recent research advancements on the topic of "Big Code" at ETH Zürich. The presentation covers several probabilistic models and how they are used (e.g., recurrent networks, CRFs, etc), for instance, JSNice. The talk also discusses various open questions and challenges that the community can explore further.

3.16 Estimating Types in Binaries using Predictive Modeling

Eran Yahav (Technion – Haifa, IL)

Reverse engineering is an important tool in mitigating vulnerabilities in binaries. As a lot of software is developed in object-oriented languages, reverse engineering of object-oriented code is of critical importance. One of the major hurdles in reverse engineering binaries compiled from object-oriented code is the use of dynamic dispatch. In the absence of debug information, any dynamic dispatch may seem to jump to many possible targets, posing a significant challenge to a reverse engineer trying to track the program flow.

We present a novel technique that allows us to statically determine the likely targets of virtual function calls. Our technique uses object tracelets – statically constructed sequences of operations performed on an object – to capture potential runtime behaviors of the object. Our analysis automatically pre-labels some of the object tracelets by relying on instances where the type of an object is known. The resulting type-labeled tracelets are then used to train a statistical language model (SLM) for each type. We then use the resulting ensemble of SLMs over unlabeled tracelets to generate a ranking of their most likely types, from which we deduce the likely targets of dynamic dispatches. We have implemented our technique and evaluated it over real-world C++ binaries. Our evaluation shows that when there are multiple alternative targets, our approach can drastically reduce the number of targets that have to be considered by a reverse engineer.

3.17 App Mining

Andreas Zeller (Universität des Saarlandes, DE)

How do we know what makes behavior correct? In the absence of detailed specifications, one alternative could be to analyze large bodies of existing software to determine which behaviors are common and thus normal. We have mined thousands of popular Android apps from the Google Play store to determine their normal behavior with respect to their API usage and their information flows. After clustering apps by their description topics, we identify outliers in each cluster with respect to their behavior. A "weather" app that sends messages thus becomes an anomaly; likewise, a “messaging” app would not be expected to access user account data. The approach is very effective in identifying novel malware even if no malware samples are given, and Google is currently adopting the approach for its store. In the long run, we expect mined patterns of normal behavior to well complement explicit specifications.

3.18 A User-Guided Approach to Program Analysis

Xin Zhang (Georgia Institute of Technology – Atlanta, US)

Program analysis tools often produce undesirable output due to various approximations. We present an approach and a system Eugene that allows user feedback to guide such approximations towards producing the desired output. We formulate the problem of user-guided program analysis in terms of solving a combination of hard rules and soft rules: hard rules capture soundness while soft rules capture degrees of approximations and preferences of users. Our technique solves the rules using an off-the-shelf solver in a manner that is sound (satisfies all hard rules), optimal (maximally satisfies soft rules), and scales to real-world analyses and programs. We evaluate Eugene on two different analyses with labeled output on a suite of seven Java programs of size 131–198 KLOC. We also report upon a user study involving nine users who employ Eugene to guide an information-flow analysis on three Java micro-benchmarks. In our experiments, Eugene significantly reduces misclassified reports upon providing limited amounts of feedback

4 Working groups

4.1 Discussion on how to advance the research along this direction and form a community around the topic

Martin T. Vechev (ETH Zürich, CH) and Veselin Raychev (ETH Zürich, CH)

On the last day of the seminar, we dedicated an hour long discussion slot. We discussed several of these questions:

Which tasks should we solve? It was observed that no two groups were solving the exact same problem. To advance the area, the goal is to define a common set of tasks that we believe are important as a community. One comment was the danger of focusing on a single task may be limiting the research. The topic of task diversity came about and that there is no need to try and be very diverse initially. An example of a possible task is: deobfuscation, as there is also a baseline here.

What metrics should we use? Machine learning metrics sometimes make no sense here. End user metrics do matter though. A point was made about algorithmic overfitting: the Siemens benchmark suite for bug localization. Plenty of works that fine-tuned details for these benchmarks, but irrelevant in general. NLP also has experience in overfitting to datasets. A good example of a metric that advanced the area is machine translation with the BLEU score (before it, all papers did user studies). Ideally, new technique papers in the area should not have to include user studies.

What is a good venue to publish the work? A comment was made that dataset papers should appear, e.g., in EMNLP there is a notable dataset award in addition to a Best paper award. Many good conferences are already accepting works in the area.

Are there expensive to obtain datasets? A question was whether we can agree on a manually annotated dataset? Which datasets are the ones that are costly and we can share the cost.

What will be important for actual programming tools? A comment was that we may actually want to overfit if we solve the actual task requested by the user. Another comment was that the particular datasets are good for our tasks, but useless for programming language tasks. Most programs do not compile. Tasks should be evaluated on how much they save for the user. Refactoring is called rarely, code completion all the time, but how much time is spent/saved for the user?

What are good outcomes of the seminar? An idea was to create a web site where everyone can upload their data set and post a challenge. It was suggested to go with Github and pull requests as opposed to a pure Wiki. This site is now a reality and several groups have already started uploading data sets and challenges:

http://learnbigcode.github.io/

Participants

- Miltos Allamanis
University of Edinburgh, GB
- Earl Barr
University College London, GB
- Jason Breck
University of Wisconsin – Madison, US
- Swarat Chaudhuri
Rice University – Houston, US
- William W. Cohen
Carnegie Mellon University, US
- Premkumar T. Devanbu
Univ. of California – Davis, US
- Shi Han
Microsoft Research – Beijing, CN
- Kenneth Heafield
University of Edinburgh, GB
- Abram Hindle
University of Alberta – Edmonton, CA
- Suresh Jagannathan
Purdue University – West Lafayette, US
- Christopher M. Jermaine
Rice University – Houston, US
- Dongsun Kim
University of Luxembourg, LU
- Dana Movshovitz-Attias
Google Inc. – Mountain View, US
- Tien N. Nguyen
Iowa State University, US
- Sebastian Proksch
TU Darmstadt, DE
- Christopher Quirk
Microsoft Corporation – Redmond, US
- Veselin Raychev
ETH Zürich, CH
- Armando Solar-Lezama
MIT – Cambridge, US
- Charles Sutton
University of Edinburgh, GB
- Daniel Tarlow
Microsoft Research UK – Cambridge, GB
- Martin T. Vechev
ETH Zürich, CH
- Nicolas Voirol
EPFL – Lausanne, CH
- Eran Yahav
Technion – Haifa, IL
- Andreas Zeller
Universität des Saarlandes, DE
- Xin Zhang
Georgia Institute of Technology – Atlanta, US

Report from Dagstuhl Seminar 15481

Evaluation in the Crowd: Crowdsourcing and Human-Centred Experiments

Edited by

Daniel Archambault[1], Tobias Hoßfeld[2], and Helen C. Purchase[3]

1 **Swansea University, GB,** D.W.Archambault@swansea.ac.uk

2 **University of Duisburg-Essen, DE,** tobias.hossfeld@uni-due.de

3 **University of Glasgow, GB,** helen.purchase@glasgow.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15481 "Evaluation in the Crowd: Crowdsourcing and Human-Centred Experiments". Human-centred empirical evaluations play important roles in the fields of human-computer interaction, visualization, graphics, multimedia, and psychology. The advent of crowdsourcing platforms, such as Amazon Mechanical Turk or Microworkers, has provided a revolutionary methodology to conduct human-centred experiments. Through such platforms, experiments can now collect data from hundreds, even thousands, of participants from a diverse user community over a matter of weeks, greatly increasing the ease with which we can collect data as well as the power and generalizability of experimental results. However, such an experimental platform does not come without its problems: ensuring participant investment in the task, defining experimental controls, and understanding the ethics behind deploying such experiments en-masse.

The major interests of the seminar participants were focused in different working groups on (W1) Crowdsourcing Technology, (W2) Crowdsourcing Community, (W3) Crowdsourcing vs. Lab, (W4) Crowdsourcing & Visualization, (W5) Crowdsourcing & Psychology, (W6) Crowdsourcing & QoE Assessment.

Seminar November 22–27, 2015 – http://www.dagstuhl.de/15481

1998 ACM Subject Classification H.1.2 User/Machine Systems, H.4 Information Systems Applications, H.5 Information Interfaces and Presentation, H.5.1 Multimedia Information Systems, H.5.3 Group and Organization Interfaces, J.0 General Computer Applications, K.4 Computers and Society, K.4.3 Organizational Impacts

Keywords and phrases Crowdsourcing, Human Computation, Crowdsourcing Design, Mechanisms, Engineering, Practical Experience, Computer Graphics, Applied Perception, HCI, Visualization

Digital Object Identifier 10.4230/DagRep.5.11.103

Edited in cooperation with Fintan McGee (Luxembourg Institute of Science and Technology, LU, fintan.mcgee@list.lu)

Evaluation in the Crowd: Crowdsourcing and Human-Centred Experiments, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 103–126
Editors: Daniel Archambault, Tobias Hoßfeld, and Helen C. Purchase
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Daniel Archambault
Tobias Hoßfeld
Helen C. Purchase

In various areas of computer science like visualization, graphics, or multimedia, it is often required to involve the users, e.g. to measure the performance of the system with respect to users, e.g. to measure the user perceived quality or usability of a system. A popular and scientifically rigorous method for assessing this performance or subjective quality is through formal experimentation, where participants are asked to perform tasks on visual representations and their performance is measured quantitatively (often through response time and errors). For the evaluation of the user perceived quality, users are conducting some experiments with the system under investigation or are completing user surveys. Also in other scientific areas like psychology, such subjective tests and user surveys are required. One approach is to conduct such empirical evaluations in the laboratory, often with the experimenter present, allowing for the controlled collection of quantitative and qualitative data. Crowdsourcing platforms can address these limitations by providing an infrastructure for the deployment of experiments and the collection of data over diverse user populations and often allows for hundreds, sometimes even thousands, of participants to be run in parallel over one or two weeks. However, when running experiments on this platform, it is hard to ensure that participants are actively engaging with the experiment and experimental controls are difficult to implement. Often, qualitative data is difficult, if not impossible, to collect as the experimenter is not present in the room to conduct an exit survey. Finally, and most importantly, the ethics behind running such experiments require further consideration. When we post a job on a crowdsourcing platform, it is often easy to forget that people are completing the job for us on the other side of the machine.

The focus of this Dagstuhl seminar was to discuss experiences and methodological considerations when using crowdsourcing platforms to run human-centred experiments to test the effectiveness of visual representations in these fields. We primarily target members of the human-computer interaction, visualization, and applied perception research as these communities often engage in human-centred experimental methodologies to evaluate their developed technologies and have deployed such technologies on crowdsourcing platforms in the past. Also, we engaged researchers that study the technology that makes crowdsourcing possible. Finally, researchers from psychology, social science and computer science that study the crowdsourcing community participated and brought another perspective on this topic. In total, 40 researchers from 13 different countries participated in the seminar. The seminar was held over one week, and included topic talks, stimulus talks and flash ('late breaking') talks. In a 'madness' session, all participants introduced themselves in a fast-paced session within 1 minutes. The participants stated their areas of interest, their expectations from the seminar, and their view on crowdsourcing science. The major interests of the participants were focused in different working groups:

- Technology to support Crowdsourcing
- Crowdworkers and the Crowdsourcing Community
- Crowdsourcing experiments vs laboratory experiments
- The use of Crowdsourcing in Psychology research
- The use of Crowdsourcing in Visualisation research
- Using Crowdsoursing to assess Quality of Experience

Figure 1 Tag cloud of the keywords from the abstracts.

The abstracts from the different talks, as well as the summary of the working groups can be found on the seminar homepage[1] and this Dagstuhl report. Apart from the report, we will produce an edited volume of articles that will become a primer text on (1) the crowdsourcing technology and methodology, (2) a comparison between crowdsourcing and lab experiments, (3) the use of crowdsourcing for visualization, psychology, and applied perception empirical studies, and (4) the nature of crowdworkers and their work, their motivation and demographic background, as well as the relationships among people forming the crowdsourcing community.

[1] http://www.dagstuhl.de/15481/

2 Table of Contents

Executive Summary

Topic Talks (60min)

Stimulus Talks (20min): Monday

Stimulus Talks (20min): Wednesday

Stimulus Talks (20min): Thursday

3 Topic Talks (60min)

3.1 Crowdsourcing Technology

Matthias Hirth (Universität Würzburg, DE) and Michelle X. Zhou (Juji Inc. – Saratoga, US)

Abstract by Matthias Hirth

Over the past few years, Crowdsourcing has become a valuable tool for researchers to easily access a large number of people in a time and cost-effective manner. This enables new possibilities for all kind of studies involving human judgements or subjective ratings. However, even if Crowdsourcing is already widely used, still open questions remain: “How to transfer lab studies to the Crowdsourcing environment?”, “How to optimize the quality of data obtained via Crowdsourcing?”, and “How to keep the same ethical standards as in lab experiment?’ are just a few of them.

To foster the discussion on possible solutions to these questions, this talk gives an overview of the current realization of the Crowdsourcing approach in commercial platforms and the resulting limitations in terms of scientific user studies. Further, it summarizes best practices and technical solutions to overcome some of those limitations or minimalise their effects on the results of crowdsourced studies.

Discussion

After the talk of Matthias Hirth the question arose, what should actually be considers as crowdsourcing. Especially the example of reCaptcha was controversial, because the users are not aware that they are working. Similarly, the use case of Crowdsensing was discussed, because users do not actively work in this case, instead they monitor passively environmental conditions with their devices. Another discussion arose about how to monitor the surrounding conditions of the crowd workers and how they influence the workers. This includes both technical conditions, e.g., the speed of the internet connection, and the physical work place of the worker, e.g., an Internet cafe. The consensus was that some of parameters, e.g., the hardware setting of the workers, do not influence the test if it is well designed and can also be checked automatically. For assessing non-technical factors, the crowd workers need to be asked explicitly. It was pointed out that also here technical solutions might be applicable, e.g., taking pictures with the worker’s web cam. However, in this case privacy issues will arise. Another question related to the workers’ background was the language of the tasks. Matthias Hirth mentioned that in the crowdsourcing platform used for his research, the task description is almost always in English. Based on that, the questions arose about whether the quality of tasks might be affected by the language of their description. On one the one hand, different languages of task descriptions might attract different workers, on the other hand, task descriptions in the native language might avoid misunderstandings. Continuing on quality control mechanisms, different methods where discussed to test if a worker is paying attention and being attentive. Here, an important outcome of the discussion was that questions testing the consistency of rating throughout a survey have to be chosen carefully. Otherwise, users might be influenced by the questions and will answer according to what they think they are expected to, instead of being truthful. Further, workers aware of these consistency questions might also communicate them (and the expected answers) to other workers via external channels, like forums. To overcome this, it was mentioned that optional free text questions might be a good indicator for the truthfulness of a worker.

Abstract by Michelle X. Zhou

As crowdsourcing becomes a more and more popular approach in human-centred studies, it is important to get a deeper understanding of the crowd who participates in such studies. Who are they? What motivates them besides financial gain? How trustworthy are they? What kind of crowd may be best suitable for what type of studies based on their characteristics and qualities? To answer these questions, both traditional and computational psychometrics technologies may be used. In this talk, I describe a computational platform that can automatically infer a crowd worker's intrinsic traits, which can be further used to determine his/her suitability for a task (e.g., image tagging vs. critical tasks) as well as predict his/her task performance.

Discussion

Michelle Zhou's talk fostered a discussion about the reliability of psychometric tests, the reasons for different test set-ups, and their comparability. It was mentioned in the discussion that each test design is based on different underlying theories and that the tests also differ, depending on their purpose. This could include for example reliability tests or utility tests. However, even if psychometric tests normally archive a Cronbach's Alpha within 0.75 and 0.83, reliability might still not be given. Additionally, it was pointed out that in these tests, rank orders may stay the same, but actual values might change. Further it was mentioned that the test-retest reliability is also depending on the test itself and not a general property. Everyone agreed that it is very important to know if someone is willing and able to cheat/deceive while participating in crowd-sourced studies. Still, the question how to detect someone's willingness to cheat (reliably) remained open.

After this discussion, the question arose, how personality tests could be included in a user study and how much time should elapse between the test and the actual task. Here one suggestion was to first run a personality test, then the actual task and one more personality test again at the end. The final question was about the semantic technologies used to analyse the free text questions Michelle Zhou mentioned in her talk.

At the end Michelle Zhou raised two questions for the discussion during the reminder of the seminar:

1. If we have a platform that can measure crowd worker's various psychological traits, such as motivations and trustworthiness, would you use it as crowd selection criteria or additional factors for result analysis?
2. What characteristics of crowd workers you believe are important to measure for your type of tasks (e.g., visualization studies)?

3.2 Crowdsourcing & Visualization

Bongshin Lee (Microsoft Research – Redmond, US) and Rita Borgo (Swansea University, GB)

Abstract

Crowdsourced labour markets represent a powerful new paradigm for accomplishing work. Visualization relies on systematic evaluation to assess effectiveness and quality of designs and tools. Evaluation methodologies however often rely on restricted, specialised user cohorts to produce hypotheses or explanations of patterns and trends in data, which in turn might not yield widely applicable results in practice. Crowdsourcing offers an interesting alternative. Yet, asking users with varying skills, backgrounds, and motivations can as well result in speculative explanations of variable quality. Understanding the crowdsourcing phenomena, social context, and environment could therefore have significant benefits for the visualization community and beyond.

Discussion

Questions:

- How reliable is the task assessment completion measure?
- What do you measure in CS experiments?
- Is there CS not related to visualization?
- What do we understand of CS as a community?
- Are CS studies comparable to lab studies?

Important aspects highlighted through the discussion:

- Characterize visualization tasks that do not require deeper understanding.
- Characterize measures and metrics for both data collection and reporting.
- Characterize and categorize workers to be able to recall them for future experiments, some CS platforms provide this functionality.
- Characterize workers abilities e.g., visual and spatial abilities, pair-match CS samples based on demographics and abilities.
- Consider features such as language differences, e.g, how do translations differ? Compare people translating from the same to the same language.
- Important to focus on: task classification, data collection vs. conditions.
- Important outcomes of the proposed Dagstuhl Book: set of definitions, categories, and an overview of current state of the art.
- There is no reproducibility in CS studies, however lab studies have also their own limitations, for example, most experiments are not repeated.

3.3 Crowdsourcing & Psychology

Edwards, Darren J. (Swansea University, GB)

Abstract

Psychology is a broad field which encompasses cognitive, social, biological, neuroscience, behavioural, clinical, and health psychology. In computer science, areas such as visualization, or Human Computer Interaction, can often rely on psychological theory (particularly cognitive psychology) to inform experimental design and develop a priori hypotheses. Today, more studies are being carried out in computer science and psychology on on-line platforms such as Mechanical Turk and Survey Monkey. In order for this novel on-line platform being accepted by the larger psychological community, replicating studies from the laboratory to the on-line platform is important. Several studies such as the Stroop task, one-shot decision making tasks, inhibition of return studies, supervised categorisation, and even brief clinical mindfulness intervention studies have been replicated. All of these tasks are quite short, and easy to complete. So, what is not known is the reliability of the on-line platform in replicating more complex tasks which may use multi-modal cognition (e.g., a variety of spatial, sequential, decision-making components), or require a large amount of attention, memory and time to complete the task. The replication of these more complex studies are needed in order to assess the potential limitations of using on-line crowdsourced platforms. However, replications can only inform the research community so far, and they do not give any information about very novel studies which have not been replicated. For this reason it is difficult to know whether these new novel study results are an artefact of on-line user environments only, and maybe cannot be replicated in the laboratory. It is assumed that the laboratory should be the gold standard, as historically, laboratory study results have been accepted as transferable to real life scenarios. However, this assumption maybe limited, as the laboratory setting also has its own biases such as participant expectancy bias. These concerns are likely to be further highlighted in the future, as more laboratories choose to use crowdsourcing as an alternative to the laboratory.

Discussion

The discussion centred around several general topics. For example, questions were asked about the nature and structure of psychology student's education, and how it came about that participation in experiments is compulsory. Also, the limitations of this setting were discussed. The problem when using psychology students, for example, is that they often know what the experiment is about, through their knowledge about psychology. They may give the experimenter what they feel he/she wants to find. Also, given the compulsory nature of laboratory experiments in that the psychology students are forced to be participants in order to obtain course credit, raises potential problems in terms of reliability of the findings. So too, do the on-line platforms have their own problems, such as “career turkers”, who are individuals seeking to make profit, or support a living from these on-line tasks. These career turkers often have multiple accounts and work on several projects at once. So, there is an obvious loss of control of the environment in which the participant works under, which can produce greater degrees of noise in the dataset. In some situations, such as sequence learning tasks, the task results rely on a strict presentation order of items, and working on multiple screens could cause false results, which are an artefact of the environment and not the task. Other points were discussed such as whether we can categorise a platform like Survey Monkey with Mechanical Turk. Survey Monkey has a dynamic graphical user interface, and you cannot create dynamic experiments with moving images, to click on, for example, or to measure reaction time. You can however use it more than for just basic surveys, such as simple categorisation studies and one-shot decision making tasks. It also has its own participant pool of users. Finally, some discussion was made about the need for graphical user interfaces being introduced so that a greater number of psychologists could use, for example, Mechanical Turk. In conclusion, both the laboratory and the on-line platform are limited in different ways. As a way forward, greater understanding of the limitations of both environments are required in order to make appropriate predictions for studies.

3.4 Getting To Know The Crowd: The Crowdsourcing Community

David Martin (Xerox Research Centre Europe – Grenoble, FR) and Neha Gupta (University of Nottingham, GB)

Abstract

This talk focused on delivering an understanding of the crowdworkers who carry out the tasks posted on micro-tasking crowdsourcing platforms like Amazon Mechanical Turk (MTurk). We contend that the correct way to view such platforms is as labour marketplaces. We provide details on the crowdworkers through sharing quantitative and qualitative research across disciplines as varied as computer science, law and sociology. We find that workers are primarily motivated by pay but that other aspects of tasks such as learning interest, and enjoyment also count. Workers face a number of problems such as unfair treatment (e.g. rejected work, blocking), low pay, information deficit and lack/asymmetry of power. We discuss some of the challenges and responsibilities for researchers who use MTurk for various types of experiments and data services. We make some remarks on how some of the conditions may be altered such that relationships may be improved between both workers and requesters. In this way it is more likely that both parties can achieve what they want from crowdsourcing while promoting respectful relationships.

Discussion

During the talk we described how we had conducted our different studies of the US Turkers and in India. The focus was very much on Turkers – covering such matters as why they go on the forum, to share all sorts of information and to work as a community, and the fact that it has a vital role in on-boarding and learning. Without the resources like the forum, it would be much harder to earn and learn. We also discussed some of the features of good jobs: good pay, prompt pay, good communication, regular posting. We had some discussion of other platforms as a comparison. There was a focus on the particular problems of outages and infrastructure for Indian Turkers. After there was a conversation that centred around how to label and categorize different types of problematic work and workers with a general feeling that we should be careful about using words like 'malicious'. A next question was around finding the right price for work, with an understanding that very low pay was not good in any way and would attract bad behaviour but that there was no simple relationship between price and quality. Finally we discussed the nature of MTurk as a market and how this poses new design questions for computer science.

4 Stimulus Talks (20min): Monday

Scribe: *Benjamin Bach (Microsoft Research – Inria Joint Centre, FR)*

This session featured four different talks covering users engagement, and motivation to participate, their background and motivation to malicious behaviour, studies in the wild, and how to improve the effectiveness of paid microtask crowdsourcing.

4.1 Evaluating Engagement and Enjoyment

Stephen G. Kobourov (University of Arizona – Tucson, US)

Abstract

While evaluation studies in InfoVis usually involve traditional performance measurements, there has been a concerted effort to move beyond time and accuracy. Of these alternative aspects, memorability and recall of visualizations have been recently considered, but other aspects such as enjoyment and engagement are not as well explored. We discuss recent studies of these topics and possible directions for future work.

Discussion

The discussion evolved around the concept of engagement and how to measure it. A first question asked about a definition of engagement; whether that meant to attract attention in the first place, or to retain attention, once it has been attracted? The answer was that both are considered, especially for information visualization.

Is it possible to measure engagement in museum-type setting? Museums have been mentioned as a potential place to measure engagement. Measuring engagement would have to take into account the place where engagement is measured, whether it's museums, public spaces, or waiting rooms, whether it's exhibition set-ups or personal devices, whether there are distractions, and so forth.

How to quantity engagement? Possible measures to quantify engagement could involve 1) How long do people stay watching or interacting with an artefact (the infographic)?, 2) Will they come back?, and 3) Will they bring someone with them? Higher values on a higher level would suggest higher overall engagement. Generally, engagement could be quantified as a matter of selection between two choices: two (or more) possibilities given, which one is chosen, and by how many?

In his informal study, the talk author has compared two network representations: node-link diagrams and map diagrams, consisting of node-link diagrams with clusters visualized as regions of a map. Results suggest that people where more “engaged” in the map representation. A question that emerged was about hypothesis or explanations *why* people stopped more often and stare at the map representations and spend more time watching? Is it because people are more familiar with maps? Is it because maps are more unusual to represent non-spatial information?

Open questions involves what motivated people to come back? Can engagement be measured, and if so, can that happen in a crowd environment given the technical constraints as well as the fact that experimenters and participants are not co-located?

4.2 Measuring Workers' Pre-Task Interactions

Jason Jacques (University of Cambridge, GB)

Abstract

The ability to entice and engage crowd workers to participate in Human Intelligence Tasks (HITs) is critical for many human computation systems and large-scale experiments. We discuss how the conversion rate of workers – the number of potential workers aware of a task that choose to accept the task – can affect the quantity, quality, and validity of any data collected via crowdsourcing. We also contribute a tool – Turkmill – that enables requesters on Amazon Mechanical Turk to easily measure the conversion rate of HITs. We investigate how four HIT design features (value proposition, branding, quality of presentation, and intrinsic motivation) affect conversion rates. Among other things, we find that including a clear value proposition has a strong significant, positive effect on the nominal conversion rate.

Discussion

The discussion evolved around the study reported in the talk. This study tried to asses what are the factors that make a task more likely to be accepted by workers.

A first question asked *Where did the workforce numbers came from?* The respective numbers (650,000 and 825,000 for the two reported studies) came from Amazon itself, but where some years old. However it was estimated that the number of active workers on Amazon Mechanical Turk (AMT) has not changed too much since then.

Did the study followed an between or within subjects design? i.e. if some workers participated in several conditions or not. The experiment was designed as a between subjects study. The experiment was listed as a single task on AMT and used a redirected workers to assign them to their tasks (group). This method has been successfully used in previous studies.

How did you measure intrinsic motivation of participants to accept a task? One of the investigated factors that make workers accepting a task was *intrinsic motivation.* The only measure of whether a task was motivating or not was acceptance, once a worker pressed the button to accept. Workers have not been asked explicitly about their intrinsic motivation. The study reported in the talk was a replication of a previous study.

Are previous study results different from this one? Previous studies have examined whether the tested factors lead to higher task accuracy. This study measured conversion rate, i.e. how the factors influence a worker's decision to accept a task.

Did you use the same AMT account for your studies? The talk reported on several studies testing for different conditions, and which all have used the same account. This may have lead to workers participating in multiple conditions, biasing the experiment. The answer was that very little overlap was reported, especially since steadily new workers are joining the pool. Further, Turkopticon ratings remained stable throughout the study period.

Did you identify reasons for changes in conversion rate and could you assess popularity? Workers have been posting tasks and on Reddit and turker forums. However, for non US-workers, there have been no reported factors.

4.3 Emotive Visual Display: Design Through Indoor and Outdoor Citizen Science?

Sara Fabrikant (Universität Zürich, CH)

Abstract

We use increasingly dynamic and mobile map displays for every-day decision making tasks (i.e., daily commutes in congested cities), and to find solutions to and communicate about complex global environmental challenges and societal needs (i.e., global climate change). However, we still have a poor understanding on how autonomic nervous activity might influence the already limited perceptual and cognitive resources of display users, for example, in time critical situations or in dilemmatic decision-making contexts (e.g., navigation, disaster mitigation and response, search and rescue, etc.).

In my talk I aim to highlight ongoing empirical research on animated and mobile map display use in the lab and in the wild, capitalizing on ambulatory human behaviour sensing methods (i.e., eye tracking, galvanic skin response, and EEG measurements). With this collected empirical data and supported by cognitive/vision theories we are guiding the process of designing maps for salience and positive engagement, thus aiming to create usable and useful visual analytics tools.

My open questions for this workshop are whether and how we might transfer this kind of direct human psycho-physiological sensing approach to evaluate visual displays into a crowdsourcing evaluation context.

Discussion

What do you consider as expertise? The reported studies investigated people with expertise in various areas. One has to differentiate between different kinds of expertise, e.g., domain expertise in a theme, or tool expertise, etc. Specifically, in the reported lab case, it included air traffic controllers trained in making decisions about aircraft movement with so-called semi-static displays, in which aircraft positions are updated every 4 seconds.

How did you obtain good measures outdoors? In the presented outdoor study, we tested expert navigators from the Swiss Armed Forces using a mobile map shown on smart device to perform a wayfinding task in unknown territory. It is generally more difficult to obtain good eye tracking data outdoors due to many often uncontrollable factors such as, weather conditions (i.e., sunlight reflected on digital displays, traffic and noise levels, and other unforeseen potentially occurring test interruptions, etc. The reported studies therefore always involve multiple measures that need to be triangulated for the analysis. For example, participants are individually shadowed by an experimenter, within a reasonable distance.

The experimenter records participants' behaviours on video and protocols unusual events. Different base line data are recorded prior to the study (i.e., spatial ability or to calibrate psycho-physiological measures) and these are then included in the data analysis.

Is that worth the effort? Could you just ask people about their feelings? Aside from self-reports, additional data collected from participants could include several psychophysiological measures including electrodermal response, eye tracking, and EEG. One can then systematically assess self-reports from questionnaires (e.g., collected on a Likert scale or open questions, etc.) with actual performance measures, to systematically assess what people say or believe, and what they actually do and how they actually perform.

Why has eye tracking been recorded in the studies? Eye tracking has been specifically employed to systematically document overt decision making process behaviour that leads to a decision or a response. This is particularly useful in combination with predetermined research hypotheses, that is, whether participants follow a theoretically predicted response pattern, i.e., formulated in a hypothesis. Eye-tracking thus can be used as process measure to explain behavioural responses.

4.4 Give Me More! Improving the Effectiveness of Paid Microtask Crowdsourcing

Ujwal Gadiraju (Leibniz Universität Hannover, DE)

Abstract

This talk discusses two pivotal aspects that influence the effectiveness of the paid crowdsourcing paradigm: (i) task design, and (ii) crowd workers' behaviour. Leveraging the dynamics of tasks that are crowdsourced on the one hand, and accounting for the behaviour of workers on the other hand, can help in designing tasks efficiently. To improve the overall quality of results, behavioural metrics can be used to measure and counter malicious activity in crowdsourced tasks. I will review some recommended guidelines for the effective design of crowdsourced surveys based on our recent research works. We also briefly discuss methods to train crowd workers to improve their performance in different types of crowdsourced microtasks.

Discussion

In assessing malicious behaviour and rule-breaking in crowdsourcing, what is your notion of rule-breaking? The notion of rule breaking in the study was the authors' one. It meant that if task rules have been violated and results were wrong or invalid. Questions have been inserted to check for workers attention, e.g. "This is an attention check question. Please select the second option". Some answers are syntactically wrong and these are easy to detect. However, answers that are semantically wrong are harder to detect.

Did workers have to pass both tests in the study? Yes, workers had to pass both attention tests in order to successfully finish the task. Tests were installed to filter workers who did not follow the task attentively and were delivering wrong answers. Workers have nevertheless been paid, even for training only.

What where the conclusions about training and tests? Training and tests reduced malicious behaviour and lead to an increased data quality.

Where did training take place in your study designs? Explicit training takes place before the tasks. The experiment design includes two separate blocks: training and experiment. *Implicit* training is inserted within the study, alternating between training and actual task sessions.

You said, the most common malicious people where of type Fast Deceivers (FD). Do people change their behaviour during tasks? The talk reported on FDs as those workers who quickly enter *some* response without checking for correctness. This behaviour is consistent throughout the tasks, i.e. workers do not become FDs throughout the task, but start as such.

What type of malicious workers appeared during training? This was not investigated in the study.

Your study reported that 40% of workers are in some way malicious. What's the motivation for malicious behaviour? We have no feedback about this. Our study wanted to track people's behaviour.

How do reputation systems works? Workers have a score, which defines their reputation. A low reputation score prevents workers from committing to certain tasks. Yet, many CS platforms do not have reputation systems set up. On the other side, the Crowdflower platform tries to create groups of similar skilled workers to specify qualifications, independent of task completion.

Do workers differ across platforms? Yes, they do. Further research is needed to compare platforms and their worker populations. There should also be research about comparing time of the day and continent of workers.

5 Stimulus Talks (20min): Wednesday

Scribe: *Sebastian Egger (AIT Austrian Institute of Technology – Wien, AT)*

This session featured four different talks, covering a framework for delivering crowdsourced evaluation of visualization, real examples of visualization evaluations using crowdsourcing, the motivation and background of crowdsourcing workers, and personal experience of crowdsourcing results in comparison with lab based results.

5.1 A crowd-sourcing framework for automated visualization evaluation

Radu Jianu (Florida International University – Miami, US)

Abstract

Due to recent research advances, the necessary ingredients for automating the process of designing and fielding user studies of data visualizations now exist. First, prescriptive rules formalize how user studies should be designed, what performance data should be collected, and how this data should be analysed. Second, task taxonomies and benchmark data standardize tasks used in evaluations. Third, a combination of advancements in web-technologies and crowd-sourcing allow even complex interactive visualizations to be evaluated

online. The talk builds on previous and current efforts in designing GraphUnit and VisUnit, two online services for crowd-sourced evaluation of visualizations, to describe how we could ultimately design and field quantitative user studies within minutes, and move our field towards benchmark driven visualization development.

Discussion

Radu Jianu reported on the advances he and his team made to develop a platform that is able to crowdsource evaluations of different visualizations without prior knowledge of web development or Java script. The resulting discussion is summarized in bullets below.

- *Is it possible to also use 3D graphs?* Yes, it's up to you which visualizations / graphs you use.
- *You have to provided functions for the evaluations. Is it code?* Indeed you have to provide code snippets
- *Are there standardized methods (ISO, IEEE...)* Rather accepted standards not standardized by an institution.
- *Which graph models do you currently support in terms of tasks: dynamic? multivariate?*
- *Was there confusion in the instructions about the technical term 'highest degree'? Did the workers have problems to identify what 'degree of . . . ' means?* Subjects had training sessions where this was explained. Training session to introduce the terms. There is also a module for language translation available.
- *Do you have a graph model that you use for verification?* Yes, we have.
- *Is it possible to run it on my own server?* Yes, it is Open Source Software.

5.2 Crowdsourcing Experiments in Visualization Research: Data, Perception, and Cognition

Remco Chang (Tufts University – Medford, US)

Abstract

In this talk I present three types of visualization experiments that we have successfully conducted on Amazon mechanical Turk (AMT). The first is on collecting user interaction patterns in a visual search task (Finding Waldo). Second is a perceptional study on modelling the perception of correlations in a scatterplot by using Weber's Law. The last is a set of cognitive experiments on priming the user's emotion (affect) and locus of control. In these experiments we demonstrated that AMT can be used to replicate laboratory studies- We further discuss some caveats and techniques for running these experiments via crowdsourcing.

Discussion

Remco Chang reported on three studies he and his team conducted with CS. One of their main aims was to gather a large quantity of data such that the were able to apply certain modelling approaches to gather either clusters of users (which share certain personality traits) based on their behaviour or to model their task performance as a function of task complexity. The resulting discussion is summarized in bullets below.

- *Did you also use psychometric tests for identifying the personality trait?* Yes, we used standardized questionnaires.
- *How did you capture and create the "locus of control"?* It's based only on mouse movement and map moves. Eye Movement is not tracked.
- *Did you track the pixels where the clicked on?* We track everything in the whole browser window. This is also one of the lessons learned, that this is necessary.
- *Were these charts identical in colour etc. ?* Yes, when both datasets were shown in one visualization we did use the same colours for all visualizations
- *How did you control brightness in the first study?* We didn't do the first study, it's a physical principle.
- *Resolution can be the same for different screen sizes.* Sure, but over 200 subjects you equal such effects out in the end.
- *How did you assess the LOC in these studies (Priming studies)?*
- *Why did you use LOC?* This was standing out well from other principles. It is more powerful than other measures, and you can also manipulate it well

5.3 Motivation of Crowd Workers, does it matter?

Babak Naderi (TU Berlin, DE)

Abstract

In my presentation, I introduced a model explaining relation between workers' motivation, outcomes and influencing factors in crowdsourcing micro-task platform based on self determination theory (Deci, Ryan 1985). On top of that we developed the Crowdsourcing Work Motivation Scale to measure the crowd workers' motivation in domain level. Furthermore, our empirical findings on how to design Trapping questions (gold standard questions) to increase the reliability of outcomes was described. Last but not least, tools provided by Quality and Usability Lab of TU-Berlin described: Crowdee the mobile crowdsourcing platform and Turkmotion , a task rating system for crowd workers.

Discussion

Babak Naderi presented a validated questionnaire for measuring the crowdworker's motivation. Furthermore, he reported on results they gathered for designing reliability questions (noticeable and hidden (=less noticeable)) in CS systems. the results showed that the noticeable reliability questions yielded better results in terms of reliability. Therefore, they concluded that the knowledge of being observed does positively impact reliability of the crowdworkers. Finally, an overview of TU Berlin's crowdee application concluded the talk. The resulting discussion is summarized in bullets below.

- *I had another concept of intrinsic motivation. Is that question (given as an example) what you base your judgement on whether or not somebody is intrinsically motivated?* It is one of the questions we use for intrinsic motivation determination, there are more questions we use to judge for intrinsic motivation.

- *You say 'unnoticeable'. Are you sure that the people do not notice the duplication of questions?* We had 97 items and it questions were separated and very difficult to notice. But perhaps it is better to label it 'less noticeable'.
- *Do you think the length of the questionnaire played also a role? Fatigue etc.* Could be, however, the longer questionnaire would typically lead to less reliable results than. But we have shown that the 'feeling of being controlled' increases the reliability score despite the longer questionnaire.
- *Motivation is to a large extent extrinsic (payment) but engagement is effected by intrinsic factors etc.*

5.4 Crowdsourcing Multimedia Experiences: Horror Stories and Lessons Learnt

Judith Redi (TU Delft, NL)

Abstract

Crowdsourcing gives researchers the opportunity to collect subjective data quickly, in the real-world, and from a very diverse pool of users. In a long-term study on image aesthetic appeal and recognizability, we challenged the crowdsourced assessments with typical lab methodologies in order to identify and analyse the impact of crowdsourcing environment on the reliability of subjective data. We identified and conducted three types of crowdsourcing experiments that helped us perform an in-depth analysis of factors influencing reliability and reproducibility of results in uncontrolled crowdsourcing environments. We provided a set of lessons learnt for future research studies which will try to port lab-based evaluation methodologies into crowdsourcing, towards avoiding the typical pitfalls in design and analysis of crowdsourcing based experiments with users.

Discussion

Judith Redi's talk gave an overview of practical experiences from a number of CS studies conducted. An interesting insight shared, was the fact that rather minor differences in interface design between lab and CS studies lead to considerably different results. Judith concluded the talk with a number of best practices that should be considered in the design of further human centred experiments. The resulting discussion is summarized in bullets below.

- *What is recognizability?* If people can identify what the image is about.
- *Who participated in the lab experiments?* Students and friend of the students that ran the experiments.
- *Recognizability could be used well across populations. Image quality is beauty and technical image quality.*
- *Is the study interface important?* Absolutely! We asked in the 2nd study "beauty" rather than "aesthetic appeal" which seems to be better comprehensible to the subjects.

6 Stimulus Talks (20min): Thursday

Scribe: *Ina Wechsung (TU Berlin, DE)*

The session covered a wide range of topics and disciplines; the first talk by Benjamin Bach was discussing whether or not complex tasks, which are usually evaluated by domain experts, can be evaluated in the crowd. Neha Gupta presented results from an ethnographic study examining crowd workers in India. A technology-driven perspective was given by Christian Keimel, who talked about hybrid devices and the opportunities offered by such devices in the context of crowd sourcing. The fourth talk in the session was held by Brian Fisher who showed how cognitive scientists make use of experimentation in the crowd.

6.1 Bringing Network visualizations to Domain Scientists

Benjamin Bach (Microsoft Research – Inria Joint Centre, FR)

Abstract

Working with domain scientists in neuroscience and history for some years, I wonder what we can learn from current crowdsourcing practices and technology to evaluate visualizations with domain scientists in the wild. Current experiments leveraging crowdsourcing focus on small and independent tasks, performable by untrained users. If we define crowdsourcing as finding and collaborating with many people you have known before, what are the limitations of such a collaboration? Can we contact and work with domain experts? Can we evaluate complex tasks? Can we measure exploration? What is a “useful” visualization? How can experts be encouraged to participate? How does the data must be?

Discussion

After the talk it was discussed how highly specialized and complex visualizations such as EEG scans can be adapted to the crowd-sourcing domain. It was debated whether or not it is possible to decompose such complex visualization as EEG scans in order to make them “crowdsourceable”. In addition, questions arose on the definition of (domain) expertise, on how to become an expert, and on the experts which we, as researchers, want to participate in our studies. It was stated that experts may be defined as people who are dealing with the specific problem in question every day.

6.2 Turk-life in India: Supporting the work of crowdwork microtask crowdsourcing

Neha Gupta (University of Nottingham, GB)

Abstract

Previous studies on Amazon Mechanical Turk (AMT), the largest marketplace for microtasks, show that the largest population of workers on AMT is U.S. based, while the second largest is based in India. In this paper, we present insights from an ethnographic study conducted in India to introduce some of these workers or "Turkers", who they are, how they work and what turking means to them. We examine the work they do to maintain their reputations and their work-life balance. In doing this, we illustrate how MT's design practically impacts on turk-work. Understanding the "lived work" of crowdwork is a valuable first step for technology design.

Discussion

The talk raised a number of questions on how researchers as employers can improve the work conditions of Turkers and allow them to have sustainable careers. For example, it was discussed if "employer recommender systems", which are currently not part of the platform, should be integrated. It was assumed that such an internal system may support the workers in getting done their communication and interaction with each other; however, it was also mentioned that a conflict of interest may arise if the system is centralized as such an "employer recommender system" primarily reflects the worker point of view. It was concluded that in order to build better systems more research is needed; we need to get a better understanding of the crowd, and the possible differences between the Indian Turkers and the US Turkers.

6.3 Crowdsourcing and hybrid devices

Christian Keimel (IRT – München, DE)

Abstract

Hybrid devices bring together the broadcast and broadband world. They provide broadcasters for the first time a direct in-program return channel for capturing consumers' feedback, resulting in new enriched "big data". Could crowdsourcing on such devices enable an in-system optimization of content distribution and maybe even content itself by taking human perception/preferences into account? Can these devices be used to bring crowdsourcing into the lean back environment in the living room, enabling new opportunities for linear and non-linear content providers?

Discussion

It was asked if such commercial platform could possibly be used for couch crowdsourcing. While this is technologically possible, as such hybrid devices offer the same possibilities as browsers, the diffusion of such devices and the services may not be sufficient yet. Nevertheless marketing research is already carried out using hybrid devices.

6.4 Psychological perspectives on crowdsourcing

Brian D. Fisher (Simon Fraser University – Surrey, CA)

Abstract

I will describe a distributed cognition perspective on visual analytics with lab studies, field studies, and translational field experiments as methods. I will then speculate a bit about ways in which crowdsourcing methods might integrate with those more developed methods in visual analytics research.

Discussion

The discussion centred around the question which experiments or tasks are (un-)suitable for crowd sourcing. It was argued that the large sample size and the diversity of crowdsourcing studies are traded off for experimental control and that this trade off is working well for some studies (e.g. video coding) but not all (e.g. psychophysiological functions). In addition, crowdsourcing is especially suited for studies which are "too big" for the lab (such as annotations of large amount of videos). It was further discussed how collaboration and related issues such as coordination problems may be studied in the crowd.

7 Flash Talks: A summary of the spontaneous talks from Thursday

Scribe: *Daniel Archambault (Swansea University, GB)*

The flash talks at this seminar provided an opportunity for participants who were inspired by the activities during the week to present small 15 minute presentations to the group for discussion. We had four such presentations by Sheelagh Carpendale, Sebastian Egger, Tatiana von Landesberger, and Fintan McGee.

Sheelagh Carpendale began first with a presentation of an overview of her laboratory's research. In particular, she presented a summary of some of her work on tabletop displays (territory, gestures in context). She also presented work on how participants construct visualizations and a finding that participants tend to construct the visualization image first and subsequently set up the axes.

During the second half of the talk, she described a crowdsourcing study where she was trying to figure out how general people interpreted words such as user, person, researcher, participant, and others by asking them to draw images of these words. In particular, the experiment tried to figure out if the images had trends according to gender and other attributes. In the first version of the study, she asked participants to draw a person. This first study produced a number of random drawings that were not people at all. By changing the question into a two phase question asking participants to first think of a person and then draw a person sitting down, nearly all participants drew people. By preparing participants for the task, good data was collected. Discussion around this talk centred around this finding and how methods for posing crowdsourcing questions is important.

Sebastian Egger presented a study on determining effective scales for evaluating video quality for use in laboratory settings and crowdsourcing applications. In this experiment, he tried to overcome a bias seen in crowdsourcing studies whereby participants would not use the full scale to judge video quality. Rather, participants would never use the bottom end of a scale $[1, 5]$. The study found that placing unclickable anchoring elements at the maximum and minimum end of the scale encouraged the participants to use the full range.

One of the interesting parts of this talk was the novel consistency/reliability checks used in this experiment. In this work, tasks where the participant was asked to list numbers displayed on the screen in order or click on stars present in a black background helped ensure that the participant was paying attention during the task. The results of these tasks were recorded to gauge participant reliability and engagement. If participant reliability was high, the participant would be put into a second pool whereby they would be called for further experiments. Using this technique allowed the experimenters to produce a more reliable pool of participants faster than previous methods.

Tatiana von Landesberger was interested in what people did with visualization techniques in the wild. She created a system, that relies on natural language processing approaches, to visualize news queries as a graph. The approach first processed news to find relationships between actors and locations and then constructed a very large network. This network could be queried to find an appropriate subgraph that could be visualized, encompassing a topic that surrounded the keyword. The system search functionality had to be tuned as certain terms consistently appeared in all graphs (e.g. Angela Merkel) due to their high connectivity.

Once the query system was created, the authors deployed both text excerpts and graphical representations of the topics to participants in a crowdsourcing study. The study found that annotations were greater in the text condition. The experimenters explained this finding by stating that the visualization helped clarify relationships in the data and therefore elaborate annotations were not needed. In a second study, they deployed the system as part of a Halloween contest. The study did not receive a large number of entries because the topic really wasn't present in the data. However, participants instead constructed graphs of things that scared them (e.g. Vladimir Putin) or constructed faces/images from the data around this theme.

Fintan McGee presented work done during the seminar on trying to determine what crowdsourcing was in terms of human-centred experimentation. Originally it was presented as a definition, but through discussion after the short presentation, it was determined that these were more characteristics of crowdsourcing experiments. The original characteristics were:

1. The participants are selected through an open anonymous call. We do not know the identities of specific participants.
2. The experiment is distributed using an online technical platform. This platform can support participant recruitment and task deployment.

3. The experiment is administered in an uncontrolled environment with limited interaction between the participant and the experimenter.

Discussion was lively and it was determined that these were properties of crowdsourcing experiments and not a definition. The second and third properties seemed integral but many seminar participants were debating if the first property was necessary.

Participants of the Seminar

In total, 40 researchers (28 male, 12 female) from 13 different countries participated in the seminar. 25 % were young researchers who actively brought in their opinions and enriched the discussions especially in the working groups.

Participants

- Daniel Archambault
Swansea University, GB
- Benjamin Bach
Microsoft Research – Inria Joint Centre, FR
- Kathrin Ballweg
TU Darmstadt, DE
- Rita Borgo
Swansea University, GB
- Alessandro Bozzon
TU Delft, NL
- Sheelagh Carpendale
University of Calgary, CA
- Remco Chang
Tufts University – Medford, US
- Min Chen
University of Oxford, GB
- Stephan Diehl
Universität Trier, DE
- Darren J. Edwards
Swansea University, GB
- Sebastian Egger
AIT Austrian Institute of Technology – Wien, AT
- Sara Fabrikant
Universität Zürich, CH
- Brian D. Fisher
Simon Fraser Univ. – Surrey, CA
- Ujwal Gadiraju
Leibniz Univ. Hannover, DE
- Neha Gupta
University of Nottingham, GB
- Matthias Hirth
Universität Würzburg, DE
- Tobias Hoßfeld
Universität Duisburg-Essen, DE
- Jason Jacques
University of Cambridge, GB
- Radu Jianu
Florida International Univ. – Miami, US
- Christian Keimel
IRT – München, DE
- Andreas Kerren
Linnaeus University – Växjö, SE
- Stephen G. Kobourov
Univ. of Arizona – Tucson, US
- Bongshin Lee
Microsoft Res. – Redmond, US
- David Martin
Xerox Research Centre Europe – Grenoble, FR
- Andrea Mauri
Polytechnic Univ. of Milan, IT
- Fintan McGee
Luxembourg Inst. of Science & Technology, LU
- Luana Micallef
HIIT – Helsinki, FI
- Sebastian Möller
TU Berlin, DE
- Babak Naderi
TU Berlin, DE
- Martin Nöllenburg
TU Wien, AT
- Helen C. Purchase
University of Glasgow, GB
- Judith Redi
TU Delft, NL
- Peter Rodgers
University of Kent, GB
- Dietmar Saupe
Universität Konstanz, DE
- Ognjen Scekic
TU Wien, AT
- Paolo Simonetto
Romano d'Ezzelino, IT
- Tatiana von Landesberger
TU Darmstadt, DE
- Ina Wechsung
TU Berlin, DE
- Michael Wybrow
Monash Univ. – Caulfield, AU
- Michelle X. Zhou
Juji Inc. – Saratoga, US

Report of Dagstuhl Seminar 15482

Social Concepts in Self-organising Systems

Edited by

Ada Diaconescu[1], Stephen Marsh[2], Jeremy Pitt[3], Wolfgang Reif[4], and Jan-Philipp Steghöfer[5]

1 **Telecom Paris Tech, FR,** ada.diaconescu@telecom-paristech.fr
2 **UOIT – Oshawa, CA,** stephen.marsh@uoit.ca
3 **Imperial College London, GB,** j.pitt@imperial.ac.uk
4 **Universität Augsburg, DE,** reif@informatik.uni-augsburg.de
5 **Chalmers UT – Göteborg, SE,** jan-philipp.steghofer@cse.gu.se

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15482 “Social Concepts in Self-organising Systems”. The seminar brought together researchers from computer sciences (in particular from the fields of multi-agent systems and self-organisation) and from social sciences to discuss the impact of the use of social concepts in technical systems as well as the interaction between technical and social systems. In an engaging and interactive setting, the problem was illuminated from a technical as well as a philosophical and legal point of view. The talks, discussions, and working groups identified a growing body of work in the field, a number of interesting and promising research avenues, as well as a set of open issues for future investigation.

Seminar November 22–27, 2015 – http://www.dagstuhl.de/15482
1998 ACM Subject Classification H.1.2 User/Machine Systems, I.2.11 Distributed Artificial Intelligence, K.4 Computers and Society
Keywords and phrases computational justice, multi-agent sytems, norms, organic computing, reputation, self-organisation, social capital, socio-technical systems, sociologically-inspired computing, trust
Digital Object Identifier 10.4230/DagRep.5.11.127

1 Executive Summary

Jan-Philipp Steghöfer
Ada Diaconescu
Stephen Marsh
Jeremy Pitt
Wolfgang Reif

There are two exciting trends in computing that motivated this seminar. On the one hand, large-scale self-organising systems gain traction in real-world settings, e.g., in the autonomous control of the power grid or in personal transportation scenarios. On the other hand, our lives are more and more pervaded by socio-technical systems that rely on the interaction of existing, complex social systems and technical systems that in many ways mirror and form the social relationships of their users. The seminar brought together researchers from a variety of domains to discuss the technical, legal, and social issues these trends incur. One focus was how social concepts can be formalised and implemented to make technical self-organising systems more robust and efficient. The other focus was how technology shapes the social system and vice versa.

Use of Social Concepts in Self-Organising and Socio-Technical Systems

The seminar's first focus is motivated by the requirements of large-scale self-organising systems. The more such systems have to take their environment into account, the more open, and the more heterogeneous they are, the more important social concepts become [1]. If a population of agents is no longer developed, deployed, maintained, and controlled by a single company or institution, the goals of the agents no longer concur—especially, the individual sub-goals no longer necessarily imply the overall system goal. In such cases, social constructs can help encourage cooperation between the agents. The presence of norms as an explicit expression of acceptable behaviour [3], of a trust management system to encourage reciprocity [4], or of a form of computational justice to settle disputes within the system [5] are measures that have been discussed in the scientific community for these ends. In this way, the systems form a legal reality that establishes certain rules and regulations within the system. This legal reality must be in accordance with the legal system under whose jurisdiction these systems work.

The second focus is how technical systems interact with and influence existing social systems. With the increasing dependence of society on computation and on complex artificial systems, their influence on human-computer interaction, and on inter-human interaction becomes a topic of concern (see, e.g., [2]). One aspect of this is the novel challenge of managing an online identity, made necessary by the representations of human users in technical systems that are, necessarily, an abstraction of the real user. Another aspect is the increasing reliance of human users on these technical aids and the potential of negative effects on the users accompanied with this. Such effects can range from infringement of privacy, to withholding of relevant information, and even to targeted manipulation.

Results of the Seminar

The seminar was highly interactive, with a lot of time dedicated to plenum discussions. Talks were used as impulses to stimulate discussion and working groups focused on particular aspects that the participants deemed particularly important.

A number of talks addressed the implications of using social concepts in technical systems from different angles and featured insights into existing technical solutions, e.g., for computational justice, trust, and ethical behaviour, as well as the observable effects of these solutions. Likewise, incentives and how social constructs influence them was a recurring theme. The discussions following these talks addressed important issues such as the relationship between system and user values, goals and the rules designed to achieve these goals, possible attacks on socio-technical systems, quantifiable incentives, and self-determination in technical systems.

A different set of talks was aimed at understanding the way social systems and technology interact, e.g., how the social organisation of the human users is represented in the technical system and becomes evident in the interactions in that system. An overview of the interplay of legal and technical systems was also provided, with important insights into the connection between technical feasibility and legal admissibility. This set of talks encouraged discussion geared towards governance, power, and the representation of values in technical systems, as well as on how to represent existing social systems in technical systems and how both the social system and its representation evolve over time.

Based on the discussions and the input from the talks, three working groups were formed that focused on discussing different aspects of the use of social concepts in self-organising systems. Their main aims and contributions are as follows:

Understanding: A first step was to consider what the notion of "social" means in the context of the technical systems regarded in the seminar. Based on a brief literature survey, the working group determined that social means that an organisation exists, that the welfare of the individuals and the organisation is regarded, and that the relations between individuals and between individuals and the organisation are a concern. A second step was to stipulate that formalising social values leads to the individuals behaving in a way that recognises their social obligations and responsibilities. Finally, the notion of "socially-sensitive design" was introduced to denote that both the design process and the system itself must be socially sensitive.

Engineering: The main concerns were how to make different social concepts usable in technical systems, how to select the fitting social construct for a specific problem, how to measure its effectiveness, and how to combine several social concepts. The working group suggested a pattern language to express selection criteria, implementation approaches, and consequences, as well as a set of metrics that make it possible to evaluate the impact of the social concept in the technical system.

Dynamics: The discussion developed towards how the social and technical components of socio-technical systems interact with each other and how the resulting dynamic aspects influence these systems. A total of six challenges were identified, the most important of which pertains to how the interaction between different social concepts that provides "checks and balances" in social systems can be transferred to technical systems. Further problems that were discussed are conflict resolution and power distribution, as well as the influence of technical systems on society and where the responsibility for this influence lies.

Future Work

The seminar participants agreed that the topic is timely and relevant and that there are a number of open issues that need to be addressed in the future. Possible venues for future elaboration of these issues are the SASOˆST workshops[1], held annually at the IEEE Conference on Self-Adaptive and Self-Organising Systems, as well as a number of other projects currently in discussion. In particular, the organisers are discussing ways to provide an overview of the state of the art of the field as well as a research roadmap and opportunities to specifically discuss the impact self-organising and socio-technical systems will have on society.

References

1 L. Vercouter and G. Muller. L.i.a.r.: Achieving social control in open and decentralized multiagent systems. *Applied Artificial Intelligence*, 24(8):723–768, 2010.

2 J. Botev. Anonymity, immediacy and electoral delegation in socio-technical networked computer systems. In M. Horbach, editor, *Informatik 2013, 43. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik angepasst an Mensch, Organisation und Umwelt, 16.-20. September 2013, Koblenz*, volume 220 of *LNI*, page 1164. GI, 2013.

3 T. Balke, C. da Costa Pereira, F. Dignum, E. Lorini, A. Rotolo, W. Vasconcelos, and S. Villata. Norms in MAS: Definitions and related concepts. In G. Andrighetto, G. Governatori,

[1] http://sasost.isse.de/

P. Noriega, and L. van der Torre, editors, *Normative Multi-Agent Systems*, volume 4 of *Dagstuhl Follow-Ups*, pages 1–31. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013.

4 L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proc. of the 35th Hawaii International Conference on System Sciences (HICSS'02)*, pages 188–196. IEEE Computer Society Press, 2002.

5 J. Pitt, D. Busquets, and R. Riveret. Formal models of social processes: The pursuit of computational justice in self-organising multi-agent systems. In *7th IEEE International Conference on Self-Adaptive and Self-Organizing Systems, SASO 2013, Philadelphia, PA, USA, September 9-13, 2013*, pages 269–270. IEEE, 2013.

2 Table of Contents

Executive Summary

Overview of Talks

Working groups

Open problems

3 Overview of Talks

3.1 When is an Interaction . . . well just an interaction

Kirstie Bellman (Topcy House Consulting, US)

The Space Station has two and a half million moving parts, which does not include the non-moving parts, the complex launch vehicle, the ground stations, the special test equipment built to support its development and continual refinement. The coordination of components, the establishment of interfaces and communications among components and the integration of required behavior from the behavior of massively different components is a major part of the decade long development of space systems. The relationships among components can be dynamic, even semi-autonomous, but they are not social. The purpose of this talk is to discuss the need for teasing apart our concepts of integration, interfaces, collective behavior, and social processes, with an emphasis on what we gain from these different levels of integration and socially-aware collective behaviors.

*with nod to Freud in title

3.2 Society of autonomous agents & Ethics

Olivier Boissier (Ecole des Mines – St. Etienne, FR)

The increasing use of multi-agent technologies in various areas raises the necessity of designing agents and societies of agents that produce ethical behaviors in context. Considering the development of socio-technical systems where humans delegate part of their decisions to agents and where user-centric approaches are required, we investigate the dimensions to be considered in order to define systems able to exhibit ethical behaviors at runtime. We claim that besides individual reasoning mechanisms and representations at the micro/agent level, representations and mechanisms have also to be investigated at the macro/society level.

3.3 Anchoring Institutions: Intermediate Institutions and the Meta-Rule of Law

Pompeu Casanovas (Autonomus University of Barcelona, ES)

How to bridge institutions, rules, norms, apps and people to set up specific ecosystems that turn “legal”? How to regulate “legally” the information flow on the Web in order to empower people (individuals and communities) and make the balance between liberty and security? How to make “legally” effective artificial devices – electronic institutions, Rights Expression Languages (REL. . .) – on the Web of Data? This presentation tries to provide some ways to answer these questions. Law has changed its traditional meaning on the Web of Data.

It requires the construction of intermediate institutions – e.g., Semantic Web Regulatory Models (SWRM) – to anchor the reuse of ontologies, datasets, and general knowledge into specific contextual legal settings. At present, there are coordination and cohesion problems between the regulatory instruments – law, policies, standards, and ethical principles – and the scenarios set forth within the Web of Data. I furnish three different examples: (i) the regulation of CAPER (a European platform to fight organised crime); (ii) lessons learned in the making of the Catalan White Book on Mediation; (iii) Licensed Linked Data. All three can be faced as cases of Relational Law.

3.4 (Socially-inspired Technology-reinforced) (Values and Pathologies)

Ada Diaconescu (Telecom Paris Tech, FR)

This presentation aims to explore the mutual relation between social and technical systems, which are both instances of self-adapting self-organising systems. The presentation raises several questions related to the opportunities and risks that may occur in the context of socially-inspired technical systems, which are executing and interacting within a social environment. Opportunities include the social values that such systems can help achieve or reinforce, even in large-scale highly-dynamic societies (different from "traditional" ones). Challenges include the social pathologies that technical systems might import along with the social concepts and processes, and which they may reinforce and magnify. Additional challenges may occur due to the lack of sufficient understanding when modelling social concepts; and/or to the context discrepancies between the social context where they evolved initially and the technical environment where they are imported. Finally, technical systems may disturb or disrupt existing social processes by altering social structures (e.g., via new communication links) and their dynamics (e.g., via reduced time scales for communication and adaptation).

3.5 Toward incentives for self-organization in a decentralized data-sharing system

Babak Esfandiari (Carleton University – Ottawa, CA)

Joint work of Esfandiari, Babak; Davoust, Alan
Main reference A. Davoust, A. Craig, B. Esfandiari, V. Kazmierski, "P2Pedia: a peer-to-peer wiki for decentralized collaboration," Concurrency and Computation: Practice and Experience, 27(11):2778–2795, 2015.
URL http://dx.doi.org/10.1002/cpe.3420

To quote Tim Berners-Lee, "we need to re-decentralize the web". This is because the web is in the hands of central authorities who can exercise control over the contributions of users, through censorship and biases in search result rankings. They can also simply disappear, taking away the entirety of user contributions. We want to evaluate an alternative where authority is decentralized among users. For such a system to work, there needs to be incentives for users to contribute and to self-organize. Producers of documents derive their payoff from the appearance of their contribution in consumer search results. Consumers derive their payoff through the relevance of the ranking of their search results. We propose ranking metrics that, while helping consumers maximize their payoff, also incentivize their participation in the system by mirroring documents and managing their network neighborhood. The diversity of these metrics also helps make the system more resilient to trust attacks (sybil, social exploitation). We sketch a simulation that helps us validate these claims.

3.6 All Human Values are System Values

Stephen Marsh (UOIT – Oshawa, CA)

Joint work of Marsh, Stephen; Dwyer, Natasha; Basu, Anirban

Sociotechnical systems are those that embody values, of their designers and creators and users – as well as the other agents in the system. To an extent, the way in which those values are represented is less important than the recognition that they exist, can be represented, and an be used for, for example, comparison (the "Jensen Question"). In this talk I examine the importance of values such as trust, peace and forgiveness, and provide Ten Commandments, originally applied to trust management, that I hope can help in the design and actualisation of all sociotechnical systems.

The commandments were first presented in:
Stephen Marsh, Anirban Basu, Natasha Dwyer: Rendering unto Cæsar the Things That Are Cæsar's: Complex Trust Models and Human Understanding. IFIPTM 2012:191–200.

Later they were increased (to 10) in an exploration in:
Stephen Marsh, Natasha Dwyer, Anirban Basu, Tim Storer, Karen Renaud, Khalil El-Khatib, Babak Esfandiari, Sylvie Noël, and Mehmet Vefa Bicakci: Foreground Trust as a Security Paradigm: Turning Users into Strong Links, Information Security in Diverse Computing Environments, 8 pages, 2014, IGI Global.

3.7 Governance, Sustainability and Justice

Jeremy Pitt (Imperial College London, GB)

Many open comuting systems – for example grid computing, sensor or vehicular networks, or virtual organisations – face a similar problem: how to collectivise and distribute resources in the absence of a centralsied controller. On approach is to define a set of conventional, mutable and mutually-agree rules – ie a self-organising rule-oriented systems in which the rules as explicit first-class entities, typically characterised by an institution. In this talk weconsider the use of self-organisig rule-oriented systems based on formal models of Ostrom's institutional design principles and Rescher's theory of distributive justice as a basis for inclusive and sustainable alloation of common-pool resources. We identify the social concepts of governance (decision-making procedures underpinning operational, collective and constitutional choice rules), sustainability (as the underlying goal of the procedures) and justness (correctness in the outcome of those procedures).

3.8 Interactional Justice

Jeremy Pitt (Imperial College London, GB)

We present interactional justice as a way to increase the 'correctness' or 'appropriateness' of outcomes of algorithmic decision- making and deliberative processes in self-organising multi-agent systems, paving the way for their effective operation in both decentralised networks and user-centred socio-technical systems. We investigate how the social concept of interactional justice can be formalised in computational logic, to understand better principles of fairness in resource allocation, inclusivity in self-determination, and fitness of procedural rules. As a result, self-organising systems can be designed and deployed for a wide range of applications, from ad hoc networks to community energy systems, wherein qualitative (social) values are primary system requirements.

3.9 Understanding Social Organizations: A Network Perspective

Ingo Scholtes (ETH Zürich, CH)

Main reference I. Scholtes, "Understanding complex systems: When Big Data meets network science," it – Information Technology, 57(4):252–256, 2015.
URL http://dx.doi.org/10.1515/itit-2015-0012

The convergence of social and technical systems provides us with a wealth of data on the structure and dynamics of social organizations. It is tempting to utilize these data to better understand how social organizations evolve, how their structure is related to their "performance", and how the position of individuals in the emerging social fabric affects their motivation. Taking a network perspective on such questions, in this talk I will introduce recent research results obtained in the context of empirical software engineering. They demonstrate the potential of computational methods in the study of social phenomena and mechanisms. At the same time, I will highlight fallacies arising in the application of the complex networks perspective to complex socio-technical systems.

References

1 Ingo Scholtes, Pavlin Mavrodiev and Frank Schweitzer: *From Aristotle to Ringelmann: A large-scale analysis of productivity and coordination in Open Source Software projects*, to appear in Empirical Software Engineering, Springer, November 2015

2 Marcelo S. Zanetti, Ingo Scholtes, Claudio Juan Tessone and Frank Schweitzer: *Categorizing Bugs with Social Networks: A Case Study on Four Open Source Software Communities*, In Proceedings of the 35th International Conference on Software Engineering (ICSE '13), pp. 1032-1041, San Francisco, CA, USA, May 2013

3 Marcelo Serrano Zanetti, Ingo Scholtes, Claudio Juan Tessone and Frank Schweitzer: *The Rise and Fall of a Central Contributor: Centralization and Performance in the Gentoo Community*, In Proceedings of the 6th International Workshop on Cooperative and Human Aspects in Software Engineering (CHASE 2013) held at ICSE 2013, San Francisco, CA, USA, May 2013

4 Emre Sarigöl, René Pfitzner, Ingo Scholtes, Antonios Garas and Frank Schweitzer: *Predicting Scientific Success Based on Coauthorship Networks*, In EPJ Data Science, Springer, September 25 2014

5 Ingo Scholtes: *Understanding Complex Systems: When Big Data meets Network Science*, In it – Information Technology, Methods and Applications of Informatics and Information Technology, August 1 2015

3.10 A few experiences on using trust for social control

Laurent Vercouter (INSA – Saint-Étienne-du-Rouvray, FR)

The social concept of trust has been widely used to implement social control for distributed systems. It is particularly suited to develop local trust assessment algorithms to evaluate the trustworthiness of an agent's acquaintances and to use it in collective decision processes for system adaptation or reconfiguration. This talk studies trust model variations defined for different use cases: peer-to-peer, social or sensor networks. We emphasize the specific characteristics of each of these cases that justifies an adaptation of the concepts or the models to obtain a relevant mapping into a (socio-)technical system. More specifically, the integration of human users in a network or the lack of an identity management system have a major impact on the way social control is implemented.

4 Working groups

4.1 Working group on Engineering Social Concepts

Olivier Boissier (Ecole des Mines – St. Etienne, FR), Gerrit Anders (Universität Augsburg, DE), Babak Esfandiari (Carleton University – Ottawa, CA), Gauthier Picard (Ecole des Mines – St. Etienne, FR), Wolfgang Reif (Universität Augsburg, DE), and Laurent Vercouter (INSA – Saint-Étienne-du-Rouvray, FR)

The working group about *engineering social concepts for self-organizing systems* has been interested in the study of the contributions of social concepts for technical and socio-technical systems and in the way they can be mapped into these systems. The use of social concepts for self-organizing systems, and more specifically socio-technical systems raises several questions and challenges from the engineering point of view.

The need for self-organizing systems arises when the system to be designed has to live in a changing and hard to predict environment, with no centralized control, and where the participating entities are potentially autonomous (this includes humans), heterogeneous and/or unreliable. Human societies share some of these properties, and the social concepts they use present the potential for being helpful metaphors and abstractions, just as ethology and biomimetism has already provided inspiration for design paradigms. Inclusiveness is another use case. When the human is in the loop, there is a need for the system to explain itself and interact using the same concepts as the ones humans use. Privacy is an area where this need seems relevant.

There are several examples of technical or socio-technical systems in which social concepts are useful. Future power management systems are one of them. They are characterized by an increasing number of small power producers, such as biogas power plants owned by farmers, and consumers, such as individual households. To ensure the system's stability and efficiency, these entities have to play an active role in the system. This imposes several challenges: Usually, power markets define a threshold for the minimum production or consumption (hereinafter referred to as prosumption) and participants have to guarantee that they can provide a specific prosumption over a certain amount of time. Furthermore, due to the shift to renewable energy sources and due to the consumers' stochastic behavior, future prosumption is subject to uncertainty. These uncertainties not only have to be taken into account to identify reliable trading partners but also to decide about the volume of a contract, for instance. There are several other application examples in which social concepts are useful in order to deal with problems such as uncertainty, the presence of human users, or the need to have decentralized and adaptive algorithms. We can cite wireless sensor networks, desktop grid computing systems, or applications in the context of smart production and smart cities, among others.

There are many challenges when attempting to translate "soft" concepts into computing systems. The concept first needs to be observed, understood and described. Then it needs to be formalized to the extent that its definition is not ambiguous. At this stage, there is the risk of leaving out crucial aspects due to oversimplification. Next, using directly the formalization as a computable model may prove intractable, and so further simplifications, translation and application to the problem domain may even reduce the usefulness the concept further. Finally, verification and validation of a technical solution may be challenging due to the lack of proper metrics and benchmarks.

In the process of engineering these systems, we encounter recurring problems and solutions based on these social metaphors. We propose to use a pattern based approach like design patterns in classical software engineering (this approach has already been used in the multi-agent domain, see [1]). These patterns can then hopefully be incorporated into a "social toolbox" which would provide us with a framework or an API supporting built-in social primitives. The self-organizing systems built with such a toolbox should then be evaluated using benchmarks that would measure the same metrics that where guiding the choice of the social pattern. These patterns are described by four fields corresponding to (i) a problem; (ii) a solution; (iii) consequences; (iv) examples.

Besides these patterns, it is necessary to have metrics and evaluation tools to emphasize the contribution of socially-based solutions. Several kinds of metrics are needed in order to evaluate the impacts (advantages/drawbacks) of introducing such concepts and mechanisms in the system. Metrics could be organized as follows depending on their functional or non functional aspects.

- Metrics on Functional Requirements
  - Application metrics, i.e. metrics related to the application domain targeted by the system
  - Social mechanism metrics, i.e. metrics related to the evaluation of the social mechanisms built from the considered social metaphors
- Metrics on Non Functional Requirements
  - Technical metrics, e.g. scalability, tractability, resilience, time to repair (mean time to failure)

- Social metrics, i.e. user acceptance, e.g. willingness to give up autonomy for a social benefit as for instance in the consumer energy area, trustworthiness of consumers such as failure to comply leading to paying a price. Dedicated benchmarks are also required so that we can make experimental comparisons of different solutions to a given problem, and thus be able to determine the relative efficiency of different social concepts or different models inspired by a same concept. Some benchmarks have already been produced in the past, such as the ART Testbed [2] for trust models assessments. These experiences have shown that the essential aspects for an efficient benchmark are:
  - to find a common scenario simulating typical problems addressed, using the social concepts we target;
  - to define specific metrics.

At last, we can emphasize a few research challenges while applying social concepts to self-organizing systems. A first one is to be able to define a computer model from an observed social phenomena going through steps of pre-formal and formal specifications as explained in the beginning of this section. Another challenge is to define a social concept pattern language to specify solutions (patterns, relations between patterns, etc.). Then specific patterns may be defined by considering a set of social concepts and a set of self-organizing system problems to find out where and how a social concept may bring an interesting solution to a given problem.

References

1 Kendall, Elizabeth A., et al. Patterns of intelligent and mobile agents. Proceedings of the second international conference on Autonomous agents. ACM, 1998.

2 Karen Fullam, Tomas B. Klos, Guillaume Muller, Jordi Sabater, Andreas Schlosser, Zvi Topol, K. Suzanne Barber, Jeffrey S. Rosenschein, Laurent Vercouter, Marco Voss, A specification of the Agent Reputation and Trust (ART) testbed: experimentation and competition for trust in agent societies. AAMAS 2005: 512–518

4.2 Dynamics in and of Social Concepts in Self-organising Systems

Sebastian Götz (TU Dresden, DE), Nelly Bencomo (Aston University – Birmingham, GB), Pompeu Casanovas (Autonomus University of Barcelona, ES), Ada Diaconescu (Telecom Paris Tech, FR), Jan Kantert (Leibniz Universität Hannover, DE), Christian Müller-Schloer (Leibniz Universität Hannover, DE), Jan-Philipp Steghöfer (Chalmers UT – Göteborg, SE), and Leon van der Torre (University of Luxembourg, LU)

The discussion on social concepts in self-organising systems can be approached based on two central questions: (a) how to transfer concepts and approaches from social sciences to technical systems and (b) how are social, technical and socio-technical systems interrelated. While the first question was addressed in the working groups on understanding and engineering, this working group focused on the second question and tailored the discussion towards the dynamic aspects of the relation between social systems and technical systems. These dynamic aspects are evident whenever processes in one of the related systems have an impact on the other. For a discussion on terminology and the engineering process to be used, we refer the interested reader to the summaries of the two respective working groups.

The main research objective of this working group was to identify a list of challenges, which are to be addressed in socio-technical systems. These challenges are intended to fuel future research and can hopefully act as the basis for a research roadmap. The following six challenges emerged from the discussions.

The first challenge denotes the need to identify various approaches for managing the mutual influence between technical and social systems.

As technological developments become ever faster, the frequency of changes to technical systems also increases. Moreover, as technology itself becomes faster and more self-adaptive, the pace and frequency of change perceivable when interacting with modern technology are even greater. This, in turn, can raise problems for socio-technical systems, since the technical parts allow for very fast responses, while the social parts respond much slower. For example, if TripAdvisor introduces a new feature such as a social network for the users, how does that change the social system constituted by its users? While the technical change is instantaneous from the perspective of the users, the actual growth of the social network and the emergence of social interactions through the technical feature can take a considerable time, if it emerges at all. In addition, such a technical service can impact the social structure of the users, their behaviour, and their interaction with the system. Additionally, since users become more reachable anywhere, at any time, via communication technology, they may experience increasing peer-pressure for more reactivity and availability. The quality of the communication may also be impacted.

A promising first step is to think about how a technical system can be integrated within a change process of the social system (e.g., as defined in a company). For this, a change process should be specified that describes how the technical system can adapt in unison with the social system. An interesting phenomenon to discuss in the context of this challenge is the *micro-macro feedback loop*. This loop denotes that the behaviour of individuals (micro-level) has an impact on a higher-level behaviour, or property, such as the overall socio-technical system, or its environment (macro-level); in turn, this higher-level phenomenon (macro-level) also has an effect on the individuals (micro-level). To approach the aforementioned challenge, this loop of mutual effects has to be considered carefully.

The second challenge is about the dangers of importing social concepts into socio-technical systems without importing the checks and balances that are part of the complex social system. Cultural evolution has produced a set of interlinked systems that provide some level of resilience to the overall collective (or system of systems) with respect to faulty or malicious behaviours from individual participants. These interlinked systems also provide means of excluding participants temporarily and of reintegrating them later on if their behaviour improves – e.g., by applying the social concept of forgiveness. The collective *together* produces a relatively stable, fair and just social system. However, in technical systems, we use isolated concepts in order to make quick decisions, which can have dire consequences for the agents who have little possibility of recourse.

For instance, the specification and enforcement of norms within a system requires that a special-purpose agent role is defined in addition, to verify the manner in which such norm-related processes are being carried-out, and hence to make sure that the associated powers are not being abused. In many legal systems, such control is, e.g., provided by the "checks and balances" achieved through the separation of powers. Likewise, when introducing a trust and reputation system, forgiveness must become part of the system to avoid the isolation of participants. Therefore, introducing new concepts from social systems may be seen like opening a Pandora's box, since it becomes necessary to import more and more social concepts to ensure that the ones the system designers originally had in mind actually work as intended.

Since self-adaptive and self-organising systems can, theoretically, change arbitrarily, and since certain behaviours can contradict the system's objectives, or inconvenience and disrupt other systems, or humans, system behaviour generally needs to be constrained, or regulated. In normative systems for instance, norms are defined to regulate, constrain and/or attach meaning to the actions of agents within the system. Here, rules and laws are made-up and agreed-upon by convention; and special-purpose mechanisms set in place to monitor and enforce them, as well as to correct non-compliant behaviour [3]. For instance, in social systems, human agents are designated to certain roles, which gives them institutional power. They can then perform acts which create conventional, mutually-agreed facts (e.g., declaring that a couple is "married"). An important question to consider here is who is allowed (or empowered) to create these facts [4]. And, in extension "Qui costudiet ipsos custodes"? Currently, in technical systems, mechanisms for ensuring checks-and-balances and conflict-resolution are most often external to the system – e.g. relying on the "traditional" legal system for settling issues in technical systems. This approach features several limitations, including the different time-scales at which technical systems and "traditional" social systems operate, and evolve – e.g. the traditional legal system lags behind modern socio-technical systems, and hence no longer addresses resulting necessities. This disregards one of Ostrom's principles [2] which requires the availability of conflict-resolution mechanisms that are fast and efficient. Additional problems may be caused if the normative system, which is external to a targeted socio-technical system, exercises too much power and hence limits the self-regulation autonomy of the socio-technical system. This also contravenes with one of Ostrom's design principles – related to the right to self-organise – and may in turn lead to the failure to manage common resources in a sustainable way (e.g., sharing electric power in a smart grid).

Since norms, laws and rules are conventional facts, rather than physical ones, and, therefore, can be broken, important challenges will be raised concerning enforcement mechanisms and their limitations. The next two challenges aim to highlight some examples of these.

The third challenge is therefore about conflict resolution, which is one of Ostrom's principles that aims to avoid the negative runaway dynamics that lead to the tragedy of the commons [1]. In other words, there is a need for inherent complexity in the regulations that police conflict resolution. For instance, the regulations on delays and reimbursements of passengers of airlines are comparatively simple, but leave a lot of room for interpretation and conflict (e.g., when should one consider that "the airplane is late"?). Nonetheless, who is doing the conflict resolution? Is this entity fair, or trustworthy? Are the resolvers empowered to act? Do the agents accept the resolution of the conflicts?

The fourth challenge identified in this working group covers the problem of imbalanced power distribution in socio-technical systems. For example, when there is a resource allocation problem and one agent feels unfairly left out of the allocation, by the time the conflict is resolved, this agent might have "starved", especially if the conflict resolution mechanism is external and lengthy. This denotes an imbalanced distribution of power, since the party withholding the resources has power over the starved individual.

Apart from the challenges above that relate directly into how to incorporate social concepts into technical systems and the consequences this might have, a fifth one deals with accountability for these consequences, both positive and negative.

The fifth challenge is concerned with the question of who takes responsibility for the impact of socio-technical systems – e.g. upon society, the environment and so on. One option would be to only hold social scientists responsible, since this would fall within their area of expertise; while computer scientists would merely be required to think about how to actually build the technical systems. In this case, lawyers would need to regulate the systems, based on advisory input from social scientists on the human-related aspects, and from computer scientists on the technical risks and limitations.

On the other hand, it could also be that computer scientists should share part of the responsibility and aim to figure-out how their systems might impact social systems, on various time scales. The precautionary principle should not be ignored here, even if it is in fact routinely ignored in practice. In this scenario, discussions and agreement among computer scientists, social scientists and lawyers would be required.

The final challenge the working group addressed is, again, a more technical one and one that is directly related to the capabilities technical systems that interact with social systems can provide.

The sixth and final challenge is about the possibility of continuous optimisation. Due to the constant interaction between the technical and the social systems it becomes possible to continuously measure the social system, adapt the technical system to the social system very quickly and, thus, to make it more suitable for its users. This optimisation could also occur for the goals of the owners or for another specific group that has different interests from the original user group.

References

1 Garrett Hardin. *The Tragedy of the Commons*. Science 162 (3859), pp. 1243–1248, 1968.

2 Elinor Ostrom. *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press, 1990.

3 Andrew Jones and Marek Sergot. *On the characterisation of law and computer systems: The normative systems perspective*, Deontic logic in computer science: normative system specification, p. 275–307, 1993

4 Andrew Jones and Marek Sergot. *A formal characterisation of institutionalised power*. Journal of the IGPL, 4(3), pp. 429–445, 1996.

4.3 Socially-Sensitive Systems Design – Working Group on "Understanding Social Concepts in Self-Organising Systems"

Peter R. Lewis (Aston University – Birmingham, GB), Kirstie Bellman (Topcy House Consulting, US), Jean Botev (University of Luxembourg, LU), Hanno Hildmann (NEC Laboratories Europe – Heidelberg, DE), Stephen Marsh (UOIT – Oshawa, CA), Jeremy Pitt (Imperial College London, GB), Ingo Scholtes (ETH Zürich, CH), and Sven Tomforde (Universität Augsburg, DE)

4.3.1 Introduction

This Working Group addressed the question of how to understand and formalise the role of social concepts in self-organising systems. Further, to achieve systems which are aware of, or in other ways sensitive to social concepts, we will be required to be able to design social concepts in to technical systems, in a principled way. In considering systems which explicitly contain social concepts (or an awareness of them), we proposed the notion of *socially-sensitive* *systems*. We then further proposed steps towards the creation of a conceptual framework for the design of socially-sensitive systems.

In a society, individuals from all walks of life voluntarily organise themselves in groups to gain benefits which improve their quality of life. There are however issues associated with free riding, scale, and time. We argued that by appealing to socially-sensitive systems design, we can support individuals by obtaining a sufficient social position to be resilient to these issues of scale and time.

Some of the new concepts discussed in socially-sensitive systems design will include how going beyond the notion that collective behaviour is only driven by immediate goals to now have concepts and mechanisms for more enduring, value-driven group behaviour and for incorporating social mechanisms supporting human social values.

One of the key benefits of socially-sensitive systems design is therefore that it leads to systems which engage in better positioning through the increase of social potential. This implies continuous redesign of the social aspects of the system, in order to ensure the resilience of the system as a whole, in a way that generalises to unknown situations. We discussed several early scenarios for how agent based system could include concepts and mechanisms for both joint goal-behaviour and for supporting the group, leading to fewer social pathologies and more robust systems.

The proposed conceptual framework for socially-sensitive systems design has three tenets: social organisation, social values and social relations. Social organisation relates to issues of network structure and roles. Social values relate to *states that matter* to individuals, and can through social mechanisms come to matter to the group as a whole. For social relations, we build on Sztompka's [9] sociological hierarchy of social relations, which differentiate a spectrum from social behaviours through social actions, to social interactions and social relations.

In this Working Group, we proposed that socially-sensitive systems design has the potential to lead to systems in which individuals behave in a way that recognises their social obligations and responsibilities. Further, we argued that this ensures the endurance of the social aspects of the system, as well as the benefits they bring.

4.3.2 Why Design Socially-Sensitive Systems

Socially-sensitive systems and design will bring benefits in two quite different types of system. In purely technical systems, a key benefit will be derived from the notion of *better positioning through increasing social potential*. Uncertainty often exists around what individuals or a group as a whole may face, or will be required to do, especially as complexity increases. Nevertheless, entities within a system will still need to achieve certain goals, often quickly. Often in a complex system, this will require interaction with, or even perhaps cooperation of, other entities within the broader system. As one example of this, resolving a resource contention issue with a degree of immediacy may require other entities to give up claims to that resource quickly. Further, these types of challenges are typically not one-offs: entities within a system may know, or learn, that they are going to have to iterate. The other entities involved in the interaction, in our example perhaps the one who is asked to give up the resource, will be encountered again. In future uncertain scenarios, perhaps the pattern will be repeated, or the roles reversed. Given the uncertainties associated with these complex systems, operating in unfolding situations, there will be a need to be able to account for and generalise to future unknown situations such as this.

In these cases, groups of individuals use social organisation, values and relations to move themselves *as a group*, to a better position to be able to deal with these factors. In doing so, the group builds what we call *social potential*[2] Without social concepts explicitly embedded in a system and its decision-making processes, there will be a lack of primitives with which to reason about the system's *social state*, and increased social potential cannot be explicitly targeted.

One example of this in the human sphere is a football team who, while they may not be under immediate attack from the opposition, nor know what form such an attack may take, apply organisation (in the form of roles and positions), values (e.g. solidarity, respect, fairness and loyalty [7]) and relations [6] to put themselves in a better position to deal with such an attack, when it does come. On a longer timescale, such social positioning has been found not only to increase the potential for success within the specific game or even group itself, but this even generalises to future life events, such as professional success [6].

However, the rise of socio-technical systems and their weaving into the fabric of human social interaction means that such values will be ever more important. Indeed, we even venture as far as to argue that such an approach is not only desirable, it is essential to preserve the intrinsic richness and value of human social interaction, as technical systems are increasingly interwoven into our everyday lives. To do otherwise, would be to degrade the quality of the human experience in a socio-technical world.

4.3.2.1 Issues With Existing Approaches

We anticipate that it will be useful to more formally understand the impact of different forms of organisation, and the role of various forms of social relation, in socio-technical systems. However, we identify a major issue with existing approaches being that of how to incorporate and manage the value we, as humans, associate with social aspects. In many traditional approaches, for example in much of multi-agent systems, an approach is generally taken whereby direct (often numerical) comparability of alternatives may be assumed, based on a universal commodification of such values.

We argue that this approach is, in general, insufficient to capture human social values such as obligation, empathy, peace and justice. Nevertheless these are the things that often truly matter the most to humans and human society. What is needed is a proper set of theories for how to actualise these things in computational systems. Such theories will be essential to realise the vision of socially-sensitive systems and design.

Our aim is to produce computational formalisations of these social values, enabling them to be explicitly represented within the technical side of socio-technical systems. In doing so, we provide some of the tools required to ensure that such systems also uphold these social values.

4.3.3 A Conceptual Framework for Socially-Sensitive Systems Design

When we talk of systems or a design process being *socially sensitive*, we mean specifically that a system is both *socially aware* and that it is *socially active.* Social awareness implies that the system can observe social aspects of its environment and interactions within in, and conceptualise these, in order to reason about social aspects. Further, *socially active* implies that a system does not simply observe and think, but based on its conceptualisations, acts in a way that is congruent with them, and its own social principles.

[2] This is not the same as social potential identified in [8], but is more closely related to social capital, in the spirit of how Fukuyama [5] sees it.

We proposed a framework for reasoning about socially-sensitive systems and their design. This is built upon three tenets:

1. The group's social organisation, including the network structure, individuals' roles and perhaps rank within it.
2. The group's social values, specifically preferences associated with states of the group and its individuals.
3. The group's social relations, both in terms of type and structure of relations.

4.3.3.1 Social Organisation

Organisation is perhaps the most familiar feature of the social nature of technical systems. It is concerned with network structure, roles of individuals and sub-groups within the system, and other features such as rank.

There is now a substantial literature on the organisation and self-organisation of (socio-)technical systems, and these aspects will underlie many of the other social concepts which a system may explicitly possess. Indeed, the organisation might be thought of as the platform, or set of constraints, upon which social relations play out, and social values are observed and propagated.

4.3.3.2 Social Values

Unpacking the notion of social values, we can relate the behaviour of individuals within a collective to obligations and responsibilities. Agents no longer care only for the goals of the group, but also for the members of the group themselves, in terms of their values for how they care for each other. In general, we can consider social values to be descriptions of *states that matter* to individuals, and can through social mechanisms come to matter to the group as a whole. More concretely, we might consider:

- States of a group that matter,
- States of an individual that matter,
- States of a relationship that matter.

States that matter to individuals can matter to groups too. This may be realised through a variety of mechanisms, such as collective decision-making and aggregation.

As discussed above, some types of value properties will be economic, insofar that they relate to things that can be readily quantified, or at least compared, without losing their primary essence. But other value properties, those that we call *welfare values*, describe qualities of the system that are less readily quantifiable or comparable. These express preferences concerning the ways in which thing are done, in accordance with what matters to individuals and groups. One potential way for formalising this notion is through the use of meta-goals, or constraints over meta-goals.

4.3.3.3 Social Relations

Weber [10] claimed that an action is 'social' if the acting individual takes account of the behaviour of others and is thereby oriented in its course. Further, Sztompka proposed [9] a hierarchy of social interactions and relations, shown in Table 1. The hierarchy makes clear how many social concepts already familiar to computer scientists relate to each other. For example, *social behaviour* is commonly analysed in ant-based systems (e.g. [2, 3]), where actions (e.g. leaving pheromone) are done for the benefit of other ants, and in doing so for the benefit of the colony as a whole. Similarly, *regular interactions* describe the kinds of

Table 1 Sztompka's Sociological Hierarchy of Social Relations.

Type	**Requires**
Behaviour	Physical movement
Action	Meaning
Social Behaviour	Directed towards others
Social Action	Await response
Social Contact	Unique / rare interaction
Social Interaction	Interactions
Repeated Interaction	Accidental, not planned, but repeated interaction
Regular Interaction	Regularity
Regulated Interaction	Interactions described by law, custom or tradition
Social Relation	A scheme of social interactions

interactions occurring in repeated games such as the iterated prisoners' dilemma [1]. Here, knowledge of future interactions with the same opponent is crucial to determining future behaviour.

However, Sztompka's hierarchy demonstrates that there are many forms of social interaction, which require varying forms of knowledge (concerning both oneself, other individuals, and the environment) as well as cognitive capabilities. In developing this tenet of our framework, our intention is to map Sztompka's framework to computational systems, and extend as necessary.

4.3.3.4 Relationship to Value-Sensitive Design

As is clear from the presented framework, in addition to a sensitivity to organisation and relations, socially-sensitive systems must be sensitive to values. Of course, this implies the practice of approaches such as value-sensitive design [4]. But socially-sensitive systems go beyond this, not just being designed by designers (or co-designed) in accordance with social values. The systems themselves will also be sensitive to such values. Thus, a form of socially-sensitive meta-design is needed. Indeed, due to additional complexities present in socio-technical self-organising systems, such as unexpected dynamics and continuous reorganisation, we may even require socially-sensitive self-design, as the system plays an active role in its own design, on a continuous basis, in accordance with social values that it itself promotes.

4.3.4 Conclusions

In summary, we propose that for socio-technical systems to possess, be aware of, and act in accordance with social concepts, these social concepts will need to be formalised and made explicit. Further, we argue that they will need to be designed in. We describe such systems and their design process as *socially-sensitive systems design*. A key benefit of socially-sensitive systems will be better positioning, through increased social potential.

While there is much work needed to formalise and realise the notion of socially-sensitive systems design, it is clear at this stage that any list of requirements for socially-sensitive systems will include at the very least:

- groups,
- awareness of others,
- directed behaviour, and
- cognition.

In this Working Group, we proposed a framework for the socially-sensitive systems design, based on three core tenets of social organisation, social values and social relations.

Ultimately, we anticipate the benefit to technical systems to include increased robustness, increased empathy with humans, a reduction of pathologies of digital communities. Further, there is also the potential for insights gained in building and using socially-sensitive systems to impact on our understanding of human society itself. As a result of this understanding feeding back into social science, we expect that we can better support human beings in society at large.

In continuing the work from this Working Group, we plan to further develop the conceptual framework which we have sketched here. This will first include mapping and possibly extending Sztompka's social relations hierarchy for computational systems. Second, we will relate existing work on individual computational values (e.g. trust) to the framework. Third, we will look to formalise other social concepts, which are not yet, or only partially explicitly present in technical systems. These include those things that often really matter to humans, such as obligation, justice and peace.

References

1 R. Axelrod. The evolution of strategies in the iterated prisoner's dilemma. In L. Davis, editor, *Genetic algorithms and simulated annealing*, pages 32–41. Pittman, London, 1987.

2 Marco Dorigo. *Optimization, Learning and Natural Algorithms*. PhD thesis, Politecnico di Milano, Milan, Italy, 1992.

3 Lukas Esterle, Peter R. Lewis, Xin Yao, and Bernhard Rinner. Socio-economic vision graph generation and handover in distributed smart camera networks. *ACM Transactions on Sensor Networks*, 10(2), 2014.

4 Batya Friedman and Peter H. Kahn Value sensitive design: Theory and methods. Technical Report 02-12-01, Deptartment Of Computer Science and Engineering, University of Washington, 2002.

5 Francis Fukuyama. *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, 1996.

6 A Huggins and S Randell. The contribution of sports to gender equality and women's empowerment. In *Proceedings of the International Conference on Gender equity on Sports for Global Change*, 2007.

7 Esther Rutten, Geert-Jan Stams, Gert Biesta, Carlo Schuengel, Evelien Dirks, and Jan Hoeksma. The contribution of organized youth sport to antisocial and prosocial behavior in adolescent athletes. *Journal of Youth and Adolescence*, 36(3):255–264, 2007.

8 John H. Reif and Hongyan Wang. Social potential fields: A distributed behavioral control for autonomous robots. *Robotics and Autonomous Systems*, 27(3):171–194, 1999.

9 Piotr Sztompka. *Socjologia*. Znak, 2002.

10 Max Weber. *The Nature of Social Action*. Cambridge University Press, 1978.

5 Open problems

5.1 Your Cheating Cat!

Stephen Marsh (UOIT – Oshawa, CA)

Main reference The talk is based on a corpus of the work I've done previously.

Technical Systems are becoming more opaque, for various reasons, some valid an some less so. However, for we (human's and otherwise) who are using or are affected by them, this is an issue. I'm beginning to explore ways to mitigate this through the various strengths of the technologies themselves. This wee ideas talk will explore the problem, think about these strengths, and try to spark discussion around our options and possible work, including in ethics, monitoring and morality.

5.2 The concept of self-reorganization

Gauthier Picard (Ecole des Mines – St. Etienne, FR)

Joint work of Picard, G., Persson, C., Boissier, O., and Ramparany, F.
Main reference G. Picard, C. Persson, O. Boissier, F. Ramparany, "Multi-agent self-organization and reorganization to adapt M2M infrastructures," in Proc. of the 9th IEEE Int'l Conf. on Self-Adaptive and Self-Organizing Systems (SASO'15), pp. 91–100, IEEE CS, 2015.
URL http://dx.doi.org/10.1109/SASO.2015.17

Designing and monitoring complex systems raise major challenges, due to the multitude of heterogeneous components, which have their own internal dynamics, while in interaction with a highly dynamic and uncertain environments. In such cases, centralised management is not realistic and predefined system behaviors lead to obsolescence. So, how to equip systems with bounded autonomous adaptation capabilities to handle these complexities and dynamics? (By bounded, we mean we want to keep control on the system, by constraining its behaviour.) From a multi-agent engineering perspective, we translate this question into "How to set up multi-agent organisation adaptation process enabling the emergence of consistent and desired behaviours to develop adaptive systems?"

We propose to join forces coming from self-organisation approaches (e.g. swarms) which exhibits important adaptiveness capabilities, and reorganisation (e.g. organisational multiagent systems) making explicit the structure, the functionalities and the constraints on the organisation. This join approach is coined "self-reorganisation". Based on multiagent programming approach JaCaMo, we share our experience on the implementation of self-reorganising systems in smart city, ambient intelligence and trust management application fields.

5.3 Why would anybody want to change?

Jan-Philipp Steghöfer (Chalmers UT – Göteborg, SE)

Change management theories recognise that organisational inertia and individual resistance are detriments to change. In this talk, I briefly describe these notions and point out the similarities to self-organising systems and their potential relevance for socio-technical systems in particular. Finally, I pose the question of whether we must consider the ability and willingness to change in agent-based systems and what the impact on self-organisation will be.

References

1 B. Burnes, "Kurt Lewin and the Planned Approach to Change: A Re-appraisal," Journal of Management studies, vol. 41, no. 6, pp. 977–1002, 2004.

2 K. E. Weick and R. E. Quinn, "Organizational Change and Development," Annual Review of Psychology, vol. 50, no. 1, pp. 361–386, 1999.

3 J. P. Kotter, "Leading Change: Why Transformation Efforts Fail," Harvard Business Review, vol. 73, no. 2, pp. 59–67, 1995.

4 Katherine J. Klein and Joann S. Sorra, "The Challenge of Innovation Implementation," The Academy of Management Review, 21(4):1055–1080, 1996.

5 Kurt Lewin, "Frontiers in Group Dynamics: Concept, Method and Reality in Social Sciences: Social Equilibria and Social Change", Human Relation, vol. 1, no. 36, 1947

Participants

- Gerrit Anders
  Universität Augsburg, DE
- Kirstie Bellman
  Topcy House Consulting, US
- Nelly Bencomo
  Aston Univ. – Birmingham, GB
- Olivier Boissier
  Ecole des Mines –
  St. Etienne, FR
- Jean Botev
  University of Luxembourg, LU
- Pompeu Casanovas
  Autonomus University of
  Barcelona, ES
- Ada Diaconescu
  Telecom Paris Tech, FR
- Babak Esfandiari
  Carleton Univ. – Ottawa, CA
- Sebastian Götz
  TU Dresden, DE
- Hanno Hildmann
  NEC Laboratories Europe –
  Heidelberg, DE
- Jan Kantert
  Leibniz Univ. Hannover, DE
- Peter R. Lewis
  Aston Univ. – Birmingham, GB
- Stephen Marsh
  UOIT – Oshawa, CA
- Christian Müller-Schloer
  Leibniz Univ. Hannover, DE
- Gauthier Picard
  Ecole des Mines –
  St. Etienne, FR
- Jeremy Pitt
  Imperial College London, GB
- Wolfgang Reif
  Universität Augsburg, DE
- Ingo Scholtes
  ETH Zürich, CH
- Jan-Philipp Steghöfer
  Chalmers UT – Göteborg, SE
- Sven Tomforde
  Universität Augsburg, DE
- Leon van der Torre
  University of Luxembourg, LU
- Laurent Vercouter
  INSA –
  Saint-Étienne-du-Rouvray, FR

Report of Dagstuhl Seminar 15491

Approximate and Probabilistic Computing: Design, Coding, Verification

Edited by

Antonio Filieri[1], Marta Kwiatkowska[2], Sasa Misailovic[3], and Todd Mytkowicz[4]

1 **Imperial College London, GB,** a.filieri@imperial.ac.uk
2 **University of Oxford, GB,** marta.kwiatkowska@cs.ox.ac.uk
3 **University of Illinois at Urbana-Champagin, US,** misailo@illinois.edu
4 **Microsoft Corporation – Redmond, US,** toddm@microsoft.com

Abstract

Computing has entered the era of *approximation*, in which hardware and software generate and reason about estimates. Navigation applications turn maps and location estimates from hardware GPS sensors into driving directions; speech recognition turns an analog signal into a likely sentence; search turns queries into information; network protocols deliver unreliable messages; and recent advances promise that approximate hardware and software will trade result quality for energy efficiency. Millions of people already use software which computes with and reasons about approximate/probabilistic data daily. These complex systems require sophisticated algorithms to deliver *accurate* answers quickly, at scale, and with energy efficiency, and approximation is often the only way to meet these competing goals.

Despite their ubiquity, economic significance, and societal impact, building such applications is difficult and requires expertise across the system stack, in addition to statistics and application-specific domain knowledge. Non-expert developers need tools and expertise to help them design, code, and verify these complex systems.

The aim of this seminar was to bring together academic and industrial researchers from the areas of probabilistic model checking, quantitative software analysis, probabilistic programming, and approximate computing to share their recent progress, identify challenges in computing with estimates, and foster collaboration with the goal of helping non-expert developers design, code, and verify modern approximate and probabilistic systems.

Seminar November 29–4, 2015 – http://www.dagstuhl.de/15491

1998 ACM Subject Classification D.2.4 Software/Program Verification, D.3.3 Language Constructs and Features, F.3.2 Semantics of Programming Languages, F.1.2 Modes of Computation

Keywords and phrases approximation, model checking, performance, probability, program analysis, systems, verification

Digital Object Identifier 10.4230/DagRep.5.11.151

Edited in cooperation with Sara Achour

Approximate and Probabilistic Computing: Design, Coding, Verification, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 151–179
Editors: Antonio Filieri, Marta Kwiatkowska, Sasa Misailovic, and Todd Mytkowicz
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Antonio Filieri
Marta Kwiatkowska
Sasa Misailovic
Todd Mytkowicz

Uncertainty and approximation are becoming first class concepts in software design and development. Many application domains, including biology, multimedia processing, finance, engineering, and social sciences, need software to formalize and study intrinsically uncertain phenomena. Furthermore, the ubiquity of software, especially driven by the Internet and mobility – such as driving applications that estimate routes, speech processing applications that estimate most likely sentences, or fitness applications that estimate heart-rate – require software engineers to design their applications taking into account unpredictable and volatile operational conditions, and noisy data, despite the limited support provided by current unintuitive design and quality assurance methodologies. Finally, the hardware community is designing devices that trade result accuracy for computational efficiency and energy saving, providing only probabilistic guarantees on the correctness of the computed results.

Several research communities are independently investigating methodologies and techniques to model, analyze, and manage uncertainty in and through software systems. These areas include (1) probabilistic model checking, (2) quantitative software analysis, (3) probabilistic programming, and (4) approximate computing. However, despite the substantial overlap of interests, researchers from different communities rarely have the opportunity to meet at conferences typically tailored to single specific areas. Therefore, we organized this seminar as a forum for industrial and academic researchers from these areas to share their recent ideas, identify the main research challenges and future directions, and explore collaborative research opportunities on problems that span across the boundaries of the individual areas.

This report presents a review of each of the main areas covered by the seminar and summarizes the discussions and conclusions of the participants.

Acknowledgements. The organizers would like to express their gratitude to the participants and the Schloss Dagstuhl team for a productive and exciting seminar. We thank Prof. Martin Rinard for his support and contribution to the organization of the seminar. We thank Sara Achour for her help with preparing this report.

2 Table of Contents

Executive Summary

Research Areas

Overview of Talks

3 Research Areas

3.1 Probabilistic Model Checking

Probabilistic modelling is widely used in the design and analysis of computer systems, and has been rapidly gaining in importance in recent years. Traditionally, models such as Markov chains have been used to analyse system performance, where typically queuing theory is applied to obtain quantitative characteristics. Probability is also needed to quantify unreliable or unpredictable behaviour, for example in fault-tolerant systems and communication protocols, where properties such as component failure and packet loss can be described probabilistically. Probabilistic models with nondeterminism, e.g., Markov decision processes, are employed for modelling of distributed co-ordination protocols which use randomisation as a symmetry breaker, in wireless medium-access control, and probabilistic routing in security and anonymity protocols. More generally, Markovian models are useful to support decision making, for example in economics, operations research, planning and robotics, to optimise a certain goal function.

Probabilistic model checking [66, 29, 5, 27] is an automatic procedure for establishing if a desired property holds in a probabilistic system model. Conventional model checkers input a description of a model, representing a state-transition system, and a specification, typically a formula in some temporal logic, and return "yes" or "no", indicating whether or not the model satisfies the specification. In the case of probabilistic model checking, the models are probabilistic (typically variants of Markov chains), in the sense that they encode the probability of making a transition between states instead of simply the existence of such a transition. A probability space induced on the system behaviours enables the calculation of likelihood of the occurrence of certain events during the execution of the system. This in turn allows one to make quantitative statements about the system [37], in addition to the qualitative statements made by conventional model checking. Probabilities are captured via probabilistic operators that extend conventional (timed or untimed) temporal logics, affording the expression of probabilistic specifications such as minimising the probability of a security attack, reliability of a nanotechnology design, and ensuring that expected energy usage of the protocol is below a specified bound.

Probabilistic model checking combines graph-theoretic analysis, drawn from conventional model checking, together with probabilistic analysis. The latter involves numerical computation, such as solving linear equations or linear programming problems, which for scalability reasons is typically implemented using iterative methods in symbolic data structures [3, 36]. This lends itself to approximate computation, where one can trade off accuracy for speed by terminating the computation early. The models are described in high-level modelling notations, or can be extracted from, e.g., C programs extended with random assignment [34]. An alternative approach, called approximate or statistical model checking [30, 68, 67], is based on simulating execution runs and applying statistical techniques such as hypothesis testing to estimate the probability or expectation of some event holding. However, no inference on data is currently combined with probabilistic model checking techniques, which focus on system dynamics. An important new direction is synthesis, which aims to construct a model that is guaranteed to satisfy a given probabilistic specification. Recently formulated and implemented simpler variants of this problem include parameter synthesis [14, 16], which finds optimal parameter values that satisfy the property and for model repair, and controller/strategy synthesis [17], with which one can generate correct-by-construction controllers from specifications.

Probabilistic model checking algorithms were proposed in the 1980s [66, 15], but it was not until early 2000s when the first industrially-relevant tools were released, notably

PRISM [38] and MRMC [33]. PRISM, in particular, is based on symbolic techniques that provide compact storage for probabilistic models and ensure efficiency of (approximate) computation of the probability. In [9], the performance of PRISM was recently improved by incorporating machine learning, with which one can obtain guarantees on accuracy while exploring only a portion of the state space. PRISM supports five probabilistic models, including probabilistic timed automata and stochastic games, for both verification and strategy synthesis. Applications of probabilistic model checking using PRISM have spanned multiple fields, from wireless protocols and source code analysis of Linux networking utilities, through debugging DNA computing designs, to smart energy grids and strategy synthesis for autonomous urban driving. The software technology underpinning probabilistic model checking has matured; it has been applied to analyse the reliability of NAND gates design, detecting a bug in an analytical model, and is being adopted, for example, in software engineering and resource management of cloud computing systems.

3.2 Quantitative Program Analysis

Probabilistic model checking developed a set of theories, algorithms, and tools aimed at verifying the properties of a variety of stochastic models. However, their applications to software engineering is mostly limited to early stages of development, where design models are translated in a more or less automatic way to corresponding stochastic models. These semantic views on the software to-be are valuable decision support systems for designers that can quantitatively evaluate the impact of their choices, especially with respect to nonfunctional requirements such as reliability or performance. However, design models are hard to keep consistent with implementation, where code artifacts are in general only partially compliant with their intended design. To mitigate this inconsistency the three main approaches are simulation [43], profiling [28], and keeping models "alive" at runtime via continuous monitoring [20]. The goal of these techniques is to perform additional measurements on the implemented artifacts in order to update the initial design assumptions as captured by design-stage models. However, these approaches can only provide coarse grained information on the implemented software that can hardly be linked to the code.

Furthermore, the widespread use of agile development processes makes the code the central, and often unique, formal model of the program. Several reverse engineering approaches attempted to automatically extract models from the code, however the extraction of meaningful models remains an open problem [10]. Black-box analysis approaches have also been proposed [64]; though useful for overall quality assessment, these approaches do not support the localization of errors or otherwise drive the improvement of the program.

Static program analysis techniques aim at checking a variety of properties of an application starting from its source code. These properties include, for example, correctness, robustness, liveness or reachability of specific statements. However, most of these techniques cannot take advantage of the characterization of uncertainty about a program inputs or about its execution flow, providing in turn less informative true-false answers. Probabilistic analysis has to be brought at the code level to support the entire development processes, from design to code and quality assurance.

Several researchers have proposed probabilistic variants of static analysis techniques, such as data flow analyses [56, 50]. In these approaches the distributions determining the probability of following each of the edges of an execution branch are supposed to be provided by the users or are coarsely estimated by monitoring a set of program executions as in [1].

Neither of these approaches is fully satisfactory since they characterize the probability of a given branch independently from the program state when the branch occurs, limiting the precision of the resulting quantitative analysis.

Probabilistic symbolic execution (PSE) is a recent technique that can be directly applied, in combination with an input probability distribution, to compute information about the probability of executing a program path, statement, or branch or, more generally, of reaching a program state [23, 21]. This technique is an example of white-box source code analysis that relies only on program semantics to quantify program behavior, taking also into account probabilistic information about its execution environment, including its deployment environment and the interaction with users and third-party components. Among the recent PSE-based techniques, [23, 21] perform an exhaustive analysis of Java programs whose branch conditions are limited to linear numeric constraints, providing precise results but suffering from scalability issues; [6] addresses the approximate analysis of non-linear constraints; [41] deals with nondeterminism and multithreaded programs; [22] provides incremental statistical analysis with quantified confidences on the results.

3.3 Probabilistic Programming

Quantitative program analysis is focused on general programs dealing with probabilistic phenomena (e.g., unpredictable interaction with users). On the other hand, probabilistic programming makes uncertainty a first-class concept and thus enables probabilistic inference.

Probabilistic programming languages augment existing programming languages with probabilistic primitives [26]. The major goal of these languages is the efficient implementation of probabilistic inference, which combines a model (written in the probabilistic programming language) with observed evidence to *infer* a distribution over variables in the program in light of that evidence. These languages abstract the details of inference, and so see frequent use by machine learning experts when building their models. Probabilistic programming has made significant strides in democratizing probabilistic inference; they let machine learning experts encode models and then ask complicated and computationally demanding queries via probabilistic inference, of those models. While, in general probabilistic inference is `NP-Hard`, probabilistic programming languages work hard to make (potentially approximate) inference efficient for many applications of practical interest.

Probabilistic programming is a well-studied field: some probabilistic programming languages such as Church [25] are theoretically universal, in that they can perform inference on any distribution they can represent. Venture [44] extends Church to allow the programmer to determine the inference algorithm to use on each part of the model. Other probabilistic programming languages restrict the distributions they allow, to make inference more tractable and efficient. Infer.NET [45, 7] uses various approximate and exact inference engines, each of which has different restrictions. For example, its Gibbs sampling [24] engine requires the distributions of related variables to be conjugate, a very strong restriction. These restrictions often require statistical expertise to evaluate, making such algorithms inappropriate for an abstraction aimed at non-experts.

Park *et al.*[54] propose a probabilistic programming language based upon sampling functions [54] which represents distributions as sampling functions, and uses operations from the probability monad [57] to build more complex distributions. Bornholt *et al.* [8] extends this idea to treat normal imperative programs, which compute with estimates, as sampling functions, thus lowering the expertise required to write a probabilistic program.

However, Bornholt *et al.*'s approach does not yet allow full probabilistic inference, like the aforementioned probabilistic programming languages.

3.4 Approximate Computing

Many modern applications are inherently approximate. For instance, multimedia processing, machine learning, and big-data analytics applications perform approximate operations on large data sets. Applications that run on today's mobile and wearable computing devices make decisions based on data from approximate hardware components (e.g., GPS, gyroscope, or accelerometer).

Up to now, developers of approximate applications had to manually reason about accuracy, energy consumption, and timely execution. Design and implementation of these applications have often been ad-hoc – hardware and software would be developed independently of each other, and integration required significant expertise at each layer of the system stack.

Approximate computing is an emerging research area that focuses on devising systematic approaches for automating development and compilation of approximate software that runs on today's commodity and approximate hardware, or tomorrow's more exotic approximate hardware. Its goal is to (1) empower a developer with the understanding of how approximate hardware and software affect the application's accuracy results, and (2) automate the management of application's accuracy, energy consumption, and performance. To achieve this goal, approximate computing brings together researchers from software systems – programming languages and software engineering – and hardware systems – circuit design and hardware architecture.

Researchers have recently proposed a number of approximate hardware designs and software optimization techniques that trade accuracy for performance and/or energy savings:

- *Approximate Hardware Architectures.* Researchers in academia have proposed a number of hardware designs with approximate accelerators or cores [39, 19, 51], ALUs [52, 18, 51], and memories [40, 61]. Typically, these designs specify the frequency of failure of their components (e.g., an addition instruction may produce a wrong result with a small probability), and/or the magnitude of error (e.g., an addition instruction may produce a small bounded noise). Researchers in industry have also proposed novel approximate hardware components, including Qualcomm's and IBM's neuromorphic accelerators [55, 32], Intel's approximate Minerva ALU design [35], and Lyric Semiconductor's (now a part of Analog Devices) belief propagation accelerator [42].
- *Approximation-Aware Compiler Optimizations.* These transformations automatically change the semantics of programs that execute on reliable (commodity) hardware to trade the accuracy of the program's result for the improved performance and/or energy consumption [58, 49, 13, 69, 47, 59]. For instance, loop perforation is a software-only technique that modifies the program to execute fewer loop iterations and therefore make the program run faster [49]. A compiler can also automate placement of operations that execute on approximate hardware [46].
- *Approximation-Aware Programming Languages and Libraries.* Programming languages such as Eon [65], EnerJ [60], and Rely [11] expose the hardware-level approximation to the developer through specific language constructs. Libraries, such as Uncertain<T> [8], provide abstractions that encapsulate approximate data within standard object-oriented programming languages. Runtime systems, such as those in Green [2], Dynamic Knobs [31], and Paraprox [59] dynamically adapt an approximate application to maintain desired result accuracy or responsiveness.

Key challenges to adopting these and other approximation techniques include characterizing their effects on the accuracy of program results and program performance. We discuss these challenges below.

- *Modeling Uncertainty:* Uncertainty can enter computation through inputs, hardware, or emerge in computation by using probabilistic language constructs. Researchers have often modeled this uncertainty probabilistically. For instance, hardware instructions produce correct results with a specified probability, a computation specifies probabilities of executing one of several approximate function versions, or the input noise has a specific probability distribution [60, 69, 48].
- *Accuracy Analysis:* Probabilistic static program analyses compute conservative bounds on the probability of large output deviations. These analyses reason about programs that operate on approximate hardware [46], programs transformed using accuracy-aware transformations [48, 69, 13], and programs that operate on uncertain inputs [63, 62]. Sampling and sensitivity testing based dynamic program analyses estimate the probability of large output deviations by running these programs on representative inputs [58, 12, 49, 4].
- *Searching for Optimal Tradeoffs:* Approximate hardware components and program transformations induce a tradeoff space between application's accuracy and performance. Optimization techniques therefore explore the tradeoff space looking for the approximate program configurations that maximize performance or energy savings subject to constraints on the accuracy of the results. Exploration can be performed using dynamic testing [58, 49, 2, 31, 59, 53], or statically reducing computation optimization to linear or integer mathematical programming [69, 46].

References

1 Glenn Ammons and James R. Larus. Improving data-flow analysis with path profiles. In *Proceedings of the ACM SIGPLAN 1998 Conference on Programming Language Design and Implementation*, PLDI'98, pages 72–84. ACM, 1998. doi:10.1145/277650.277665.

2 Woongki Baek and Trishul M. Chilimbi. Green: A framework for supporting energy-conscious programming using controlled approximation. In *Proceedings of the 2010 ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI'10, pages 198–209, New York, NY, USA, 2010. ACM. doi:10.1145/1806596.1806620.

3 Christel Baier, Edmund M. Clarke, Vasiliki Hartonas-Garmhausen, Marta Kwiatkowska, and Mark Ryan. Symbolic model checking for probabilistic processes. In Pierpaolo Degano, Roberto Gorrieri, and Alberto Marchetti-Spaccamela, editors, *Automata, Languages and Programming*, volume 1256 of *Lecture Notes in Computer Science*, pages 430–440. Springer, 1997. doi:10.1007/3-540-63165-8_199.

4 Tao Bao, Yunhui Zheng, and Xiangyu Zhang. White box sampling in uncertain data processing enabled by program analysis. In *Proceedings of the ACM International Conference on Object Oriented Programming Systems Languages and Applications*, OOPSLA'12, pages 897–914. ACM, 2012. doi:10.1145/2384616.2384681.

5 Andrea Bianco and Luca de Alfaro. Model checking of probabilistic and nondeterministic systems. In P.S. Thiagarajan, editor, *Foundations of Software Technology and Theoretical Computer Science*, volume 1026 of *Lecture Notes in Computer Science*, pages 499–513. Springer, 1995. doi:10.1007/3-540-60692-0_70.

6 Mateus Borges, Antonio Filieri, Marcelo d'Amorim, Corina S. Păsăreanu, and Willem Visser. Compositional solution space quantification for probabilistic software analysis. In *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI'14, pages 123–132. ACM, 2014. doi:10.1145/2594291.2594329.

7 Johannes Borgström, Andrew D. Gordon, Michael Greenberg, James Margetson, and Jurgen Van Gael. Measure transformer semantics for bayesian machine learning. In Gilles Barthe, editor, *Programming Languages and Systems*, volume 6602 of *Lecture Notes in Computer Science*, pages 77–96. Springer, 2011. doi:10.1007/978-3-642-19718-5_5.

8 James Bornholt, Todd Mytkowicz, and Kathryn S. McKinley. Uncertain<t>: A first-order type for uncertain data. In *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS'14, pages 51–66, New York, NY, USA, 2014. ACM. doi:10.1145/2541940.2541958.

9 T. Brázdil, K. Chatterjee, M. Chmelík, V. Forejt, J. Křetínský, M. Kwiatkowska, D. Parker, and M. Ujma. Verification of markov decision processes using learning algorithms. In *Proc. 12th International Symposium on Automated Technology for Verification and Analysis (ATVA'14)*, LNCS. Springer, 2014. To appear.

10 Gerardo Canfora, Massimiliano Di Penta, and Luigi Cerulo. Achievements and challenges in software reverse engineering. *Commun. ACM*, 54(4):142–151, April 2011. doi:10.1145/1924421.1924451.

11 Michael Carbin, Sasa Misailovic, and Martin C. Rinard. Verifying quantitative reliability for programs that execute on unreliable hardware. In *Proceedings of the 2013 ACM SIGPLAN International Conference on Object Oriented Programming Systems Languages and Applications*, OOPSLA'13, pages 33–52, New York, NY, USA, 2013. ACM. doi:10.1145/2509136.2509546.

12 Michael Carbin and Martin C. Rinard. Automatically identifying critical input regions and code in applications. In *Proceedings of the 19th International Symposium on Software Testing and Analysis*, ISSTA'10, pages 37–48. ACM, 2010. doi:10.1145/1831708.1831713.

13 Swarat Chaudhuri, Sumit Gulwani, Roberto Lublinerman, and Sara Navidpour. Proving programs robust. In *Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering*, ESEC/FSE'11, pages 102–112, New York, NY, USA, 2011. ACM. doi:10.1145/2025113.2025131.

14 Taolue Chen, E.M. Hahn, Tingting Han, M. Kwiatkowska, Hongyang Qu, and Lijun Zhang. Model repair for markov decision processes. In *Theoretical Aspects of Software Engineering (TASE), 2013 International Symposium on*, pages 85–92, July 2013. doi:10.1109/TASE.2013.20.

15 Costas Courcoubetis and Mihalis Yannakakis. Markov decision processes and regular events. In MichaelS. Paterson, editor, *Automata, Languages and Programming*, volume 443 of *Lecture Notes in Computer Science*, pages 336–349. Springer, 1990. doi:10.1007/BFb0032043.

16 M. Diciolla, C. H. P. Kim, M. Kwiatkowska, and A. Mereacre. Synthesising optimal timing delays for timed i/o automata. In *14th International Conference on Embedded Software (EMSOFT'14)*, 2014 - to appear.

17 Klaus Drager, Vojtěch Forejt, Marta Kwiatkowska, David Parker, and Mateusz Ujma. Permissive controller synthesis for probabilistic systems. In Erika Abraham and Klaus Havelund, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 8413 of *Lecture Notes in Computer Science*, pages 531–546. Springer, 2014. doi:10.1007/978-3-642-54862-8_44.

18 Hadi Esmaeilzadeh, Adrian Sampson, Luis Ceze, and Doug Burger. Architecture support for disciplined approximate programming. In *Proceedings of the Seventeenth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS XVII, pages 301–312. ACM, 2012. doi:10.1145/2150976.2151008.

19 Hadi Esmaeilzadeh, Adrian Sampson, Luis Ceze, and Doug Burger. Neural acceleration for general-purpose approximate programs. In *Proceedings of the 2012 45th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO-45, pages 449–460. IEEE Computer Society, 2012. doi:10.1109/MICRO.2012.48.

20 Antonio Filieri, Carlo Ghezzi, and Giordano Tamburrelli. A formal approach to adaptive software: continuous assurance of non-functional requirements. *Formal Aspects of Computing*, 24(2):163–186, 2012. doi:10.1007/s00165-011-0207-2.

21 Antonio Filieri, Corina S. Păsăreanu, and Willem Visser. Reliability analysis in symbolic pathfinder. In *Proceedings of the 2013 International Conference on Software Engineering*, ICSE'13, pages 622–631. IEEE Press, 2013. doi:10.1109/ICSE.2013.6606608.

22 Antonio Filieri, Corina S. Păsăreanu, Willem Visser, and Jaco Geldenhuys. Statistical symbolic execution with informed sampling. In *Proceedings of the ACM SIGSOFT 22nd International Symposium on the Foundations of Software Engineering*, FSE'14. ACM, 2014. URL: http://goo.gl/GXxFLi.

23 Jaco Geldenhuys, Matthew B. Dwyer, and Willem Visser. Probabilistic symbolic execution. In *Proceedings of the 2012 International Symposium on Software Testing and Analysis*, ISSTA 2012, pages 166–176. ACM, 2012. doi:10.1145/2338965.2336773.

24 Stuart Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(6):721–741, Nov 1984. doi:10.1109/TPAMI.1984.4767596.

25 Noah D. Goodman, Vikash K. Mansinghka, Daniel M. Roy, Keith Bonawitz, and Joshua B. Tenenbaum. Church: A language for generative models. In *Uncertainty in Artificial Intelligence*, pages 220–229, 2008. URL: http://arxiv.org/pdf/1206.3255.

26 Andrew D. Gordon, Thomas A. Henzinger, Aditya V. Nori, and Sriram K. Rajamani. Probabilistic programming. In *Proceedings of the on Future of Software Engineering*, FOSE 2014, pages 167–181, New York, NY, USA, 2014. ACM. doi:10.1145/2593882.2593900.

27 Christel gordon and Marta Kwiatkowska. Model checking for a probabilistic branching time logic with fairness. *Distributed Computing*, 11(3):125–155, 1998. doi:10.1007/s004460050046.

28 K. Goseva-Popstojanova, M. Hamill, and R. Perugupalli. Large empirical case study of architecture-based software reliability. In *Software Reliability Engineering, 2005. ISSRE 2005. 16th IEEE International Symposium on*, pages 52–61, Nov 2005. doi:10.1109/ISSRE.2005.25.

29 Hans Hansson and Bengt Jonsson. A logic for reasoning about time and reliability. *Formal Aspects of Computing*, 6(5):512–535, 1994. doi:10.1007/BF01211866.

30 Thomas Herault, Richard Lassaigne, Frederic Magniette, and Sylvain Peyronnet. Approximate probabilistic model checking. In Bernhard Steffen and Giorgio Levi, editors, *Verification, Model Checking, and Abstract Interpretation*, volume 2937 of *Lecture Notes in Computer Science*, pages 73–84. Springer Berlin Heidelberg, 2004. doi:10.1007/978-3-540-24622-0_8.

31 Henry Hoffmann, Stelios Sidiroglou, Michael Carbin, Sasa Misailovic, Anant Agarwal, and Martin Rinard. Dynamic knobs for responsive power-aware computing. In *Proceedings of the Sixteenth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS XVI, pages 199–212, New York, NY, USA, 2011. ACM. doi:10.1145/1950365.1950390.

32 New ibm synapse chip could open era of vast neural networks. http://www-03.ibm.com/press/us/en/pressrelease/44529.wss.

33 Joost-Pieter Katoen, Ivan S. Zapreev, Ernst Moritz Hahn, Holger Hermanns, and David N. Jansen. The ins and outs of the probabilistic model checker mrmc. *Perform. Eval.*, 68(2):90–104, February 2011. doi:10.1016/j.peva.2010.04.001.

34 M. Kattenbelt, M. Kwiatkowska, G. Norman, and D. Parker. Abstraction refinement for probabilistic software. In N. Jones and M. Muller-Olm, editors, *Proc. 10th International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI'09)*, volume 5403 of *LNCS*, pages 182–197. Springer, 2009.

35 H. Kaul, M. Anders, S. Mathew, S. Hsu, A. Agarwal, F. Sheikh, R. Krishnamurthy, and S. Borkar. A 1.45ghz 52-to-162gflops/w variable-precision floating-point fused multiply-add unit with certainty tracking in 32nm cmos. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, pages 182–184, Feb 2012. doi:10.1109/ISSCC.2012.6176987.

36 M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model checking with PRISM: A hybrid approach. *International Journal on Software Tools for Technology Transfer (STTT)*, 6(2):128–142, 2004.

37 Marta Kwiatkowska. Quantitative verification: Models techniques and tools. In *Proceedings of the the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering*, ESEC-FSE'07, pages 449–458. ACM, 2007. doi:10.1145/1287624.1287688.

38 Marta Kwiatkowska, Gethin Norman, and David Parker. Prism 4.0: Verification of probabilistic real-time systems. In Ganesh Gopalakrishnan and Shaz Qadeer, editors, *Computer Aided Verification*, volume 6806 of *Lecture Notes in Computer Science*, pages 585–591. Springer, 2011. doi:10.1007/978-3-642-22110-1_47.

39 L. Leem, Hyungmin Cho, J. Bau, Q.A. Jacobson, and S. Mitra. Ersa: Error resilient system architecture for probabilistic applications. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pages 1560–1565, March 2010. doi:10.1109/DATE.2010.5457059.

40 Song Liu, Karthik Pattabiraman, Thomas Moscibroda, and Benjamin G. Zorn. Flikker: Saving dram refresh-power through critical data partitioning. In *Proceedings of the Sixteenth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS XVI, pages 213–224. ACM, 2011. doi:10.1145/1950365.1950391.

41 Kasper Luckow, Corina S. Păsăreanu, Matthew B. Dwyer, Antonio Filieri, and Willem Visser. Exact and approximate probabilistic symbolic execution for nondeterministic programs. In *Proceedings of the 29th ACM/IEEE International Conference on Automated Software Engineering*, ASE'14, pages 575–586. ACM, 2014. doi:10.1145/2642937.2643011.

42 Lyriclabs: High probability of success. http://newsoffice.mit.edu/2013/ben-vigoda-lyric-0501.

43 Michael R. Lyu, editor. *Handbook of Software Reliability Engineering.* McGraw-Hill, Inc., Hightstown, NJ, USA, 1996.

44 Vikash K. Mansinghka, Daniel Selsam, and Yura N. Perov. Venture: a higher-order probabilistic programming platform with programmable inference. *CoRR*, abs/1404.0099, 2014. URL: http://arxiv.org/abs/1404.0099.

45 T. Minka, J.M. Winn, J.P. Guiver, and D.A. Knowles. Infer.NET 2.5, 2012. Microsoft Research Cambridge. URL: http://research.microsoft.com/infernet.

46 Sasa Misailovic, Michael Carbin, Sara Achour, Zichao Qi, and Martin C. Rinard. Chisel: Reliability- and accuracy-aware optimization of approximate computational kernels. In *Proceedings of the 2014 ACM International Conference on Object Oriented Programming Systems Languages and Applications*, OOPSLA'14, pages 309–328. ACM, 2014. doi:10.1145/2660193.2660231.

47 Sasa Misailovic, Deokhwan Kim, and Martin Rinard. Parallelizing sequential programs with statistical accuracy tests. *ACM Trans. Embed. Comput. Syst.*, 12(2s):88:1–88:26, May 2013. doi:10.1145/2465787.2465790.

48 Sasa Misailovic, Daniel M. Roy, and MartinC. Rinard. Probabilistically accurate program transformations. In Eran Yahav, editor, *Static Analysis*, volume 6887 of *Lecture Notes in Computer Science*, pages 316–333. Springer, 2011. doi:10.1007/978-3-642-23702-7_24.

49 Sasa Misailovic, Stelios Sidiroglou, Henry Hoffmann, and Martin Rinard. Quality of service profiling. In *Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering*, ICSE'10, pages 25–34, New York, NY, USA, 2010. ACM. `doi:10.1145/1806799.1806808`.

50 David Monniaux. An abstract monte-carlo method for the analysis of probabilistic programs. In *Proceedings of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL'01, pages 93–101. ACM, 2001. `doi:10.1145/360204.360211`.

51 Sriram Narayanan, John Sartori, Rakesh Kumar, and Douglas L. Jones. Scalable stochastic processors. In *Proceedings of the Conference on Design, Automation and Test in Europe*, DATE'10, pages 335–338. European Design and Automation Association, 2010. `doi:10.1109/DATE.2010.5457181`.

52 K.V. Palem. Energy aware computing through probabilistic switching: a study of limits. *Computers, IEEE Transactions on*, 54(9):1123–1137, Sept 2005. `doi:10.1109/TC.2005.145`.

53 J. Park, X. Zhang, K. Ni, H. Esmaeilzadeh, and M. Naik. Expectation-oriented framework for automating approximate programming. Technical Report GT-CS-14-05, Georgia Institute of Technology, 2014. URL: https://smartech.gatech.edu/handle/1853/49755.

54 Sungwoo Park, Frank Pfenning, and Sebastian Thrun. A probabilistic language based on sampling functions. *ACM Trans. Program. Lang. Syst.*, 31(1):4:1–4:46, December 2008. `doi:10.1145/1452044.1452048`.

55 Introducing qualcomm zeroth processors: Brain-inspired computing. https://www.qualcomm.com/news/onq/2013/10/10/introducing-qualcomm-zeroth-processors-brain-inspired-computing.

56 G. Ramalingam. Data flow frequency analysis. In *Proceedings of the ACM SIGPLAN 1996 Conference on Programming Language Design and Implementation*, PLDI'96, pages 267–277. ACM, 1996. `doi:10.1145/231379.231433`.

57 Norman Ramsey and Avi Pfeffer. Stochastic lambda calculus and monads of probability distributions. In *Proceedings of the 29th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL'02, pages 154–165. ACM, 2002. `doi:10.1145/503272.503288`.

58 Martin Rinard. Probabilistic accuracy bounds for fault-tolerant computations that discard tasks. In *Proceedings of the 20th Annual International Conference on Supercomputing*, ICS'06, pages 324–334. ACM, 2006. `doi:10.1145/1183401.1183447`.

59 Mehrzad Samadi, Davoud Anoushe Jamshidi, Janghaeng Lee, and Scott Mahlke. Paraprox: Pattern-based approximation for data parallel applications. In *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS'14, pages 35–50, New York, NY, USA, 2014. ACM. `doi:10.1145/2541940.2541948`.

60 Adrian Sampson, Werner Dietl, Emily Fortuna, Danushen Gnanapragasam, Luis Ceze, and Dan Grossman. Enerj: Approximate data types for safe and general low-power computation. In *Proceedings of the 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI'11, pages 164–174. ACM, 2011. `doi:10.1145/1993498.1993518`.

61 Adrian Sampson, Jacob Nelson, Karin Strauss, and Luis Ceze. Approximate storage in solid-state memories. In *Proceedings of the 46th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO-46, pages 25–36, New York, NY, USA, 2013. ACM. `doi:10.1145/2540708.2540712`.

62 Adrian Sampson, Pavel Panchekha, Todd Mytkowicz, Kathryn S. McKinley, Dan Grossman, and Luis Ceze. Expressing and verifying probabilistic assertions. In *Proceedings of*

the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI'14, pages 112–122, New York, NY, USA, 2014. ACM. doi:10.1145/2594291.2594294.

63 Sriram Sankaranarayanan, Aleksandar Chakarov, and Sumit Gulwani. Static analysis for probabilistic programs: Inferring whole program properties from finitely many paths. In *Proceedings of the 34th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI'13, pages 447–458, New York, NY, USA, 2013. ACM. doi:10.1145/2491956.2462179.

64 Koushik Sen, Mahesh Viswanathan, and Gul Agha. Statistical model checking of black-box probabilistic systems. In Rajeev Alur and Doron A. Peled, editors, *Computer Aided Verification*, volume 3114 of *Lecture Notes in Computer Science*, pages 202–215. Springer, 2004. doi:10.1007/978-3-540-27813-9_16.

65 Jacob Sorber, Alexander Kostadinov, Matthew Garber, Matthew Brennan, Mark D. Corner, and Emery D. Berger. Eon: A language and runtime system for perpetual systems. In *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, SenSys'07, pages 161–174. ACM, 2007. doi:10.1145/1322263.1322279.

66 M.Y. Vardi. Automatic verification of probabilistic concurrent finite state programs. In *Foundations of Computer Science, 1985., 26th Annual Symposium on*, pages 327–338, Oct 1985. doi:10.1109/SFCS.1985.12.

67 HakanL.S. Younes, EdmundM. Clarke, and Paolo Zuliani. Statistical verification of probabilistic properties with unbounded until. In Jim Davies, Leila Silva, and Adenilso Simao, editors, *Formal Methods: Foundations and Applications*, volume 6527 of *Lecture Notes in Computer Science*, pages 144–160. Springer, 2011. doi:10.1007/978-3-642-19829-8_10.

68 HåkanL.S. Younes, Marta Kwiatkowska, Gethin Norman, and David Parker. Numerical vs. statistical probabilistic model checking. *International Journal on Software Tools for Technology Transfer*, 8(3):216–228, 2006. doi:10.1007/s10009-005-0187-8.

69 Zeyuan Allen Zhu, Sasa Misailovic, Jonathan A. Kelner, and Martin Rinard. Randomized accuracy-aware program transformations for efficient approximate computations. In *Proceedings of the 39th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL'12, pages 441–454. ACM, 2012. doi:10.1145/2103656.2103710.

4 Overview of Talks

4.1 Approximate computation with outlier detection in Topaz

Sara Achour (MIT – Cambridge, US)

Joint work of Achour, Sara; Rinard, Martin
Main reference S. Achour, M. C. Rinard, "Approximate computation with outlier detection in Topaz," in Proc. of the 2015 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA'15), pp. 711–730, ACM, 2015.
URL http://dx.doi.org/10.1145/2814270.2814314

We present Topaz, a new task-based language for computations that execute on approximate computing platforms that may occasionally produce arbitrarily inaccurate results. Topaz maps tasks onto the approximate hardware and integrates the generated results into the main computation. To prevent unacceptably inaccurate task results from corrupting the main computation, Topaz deploys a novel outlier detection mechanism that recognizes and precisely re-executes outlier tasks. Outlier detection enables Topaz to work effectively with approximate hardware platforms that have complex fault characteristics, including platforms with bit pattern dependent faults (in which the presence of faults may depend on values stored in adjacent memory cells). Our experimental results show that, for our set of benchmark applications, outlier detection enables Topaz to deliver acceptably accurate results (less than 1% error) on our target approximate hardware platforms. Depending on the application and the hardware platform, the overall energy savings range from 5 to 13 percent. Without outlier detection, only one of the applications produces acceptably accurate results.

4.2 Numerical Program Analysis Tools: A Wish List

David Bindel (Cornell University, US)

In my scientific computing work, I am constantly faced with different sources of error: model error, stochastic error, discretization error, error due to approximation of some difficult term, error due to termination of iterations, and error due to roundoff effects. I deal with these errors by reasoning about forward and backward errors, stability and conditioning of iterations and of problems, the role of singularities, and structural properties of the computation must be retained for meaningful results. I dream of compilers with which I can share optimizations that I know are possible (and those that will break my code) and PL tools that understand enough to help me check my error analyses. I will share some of my own preliminary work in this direction, and will make an appeal to the audience to help produce the tools I wish I knew how to write.

4.3 Optimizing Synthesis with Metasketches (for Automated Approximate Programming)

James Bornholt (University of Washington – Seattle, US)

URL http://synapse.uwplse.org/

An ideal programming model for approximate computing would apply approximations automatically, translating an exact program and a quality specification into the most efficient program that meets that specification. Program synthesis is the task of automatically generating a program that meets a given specification, and sounds like a good fit for the approximate computing problem. But existing synthesis tools rarely consider the efficiency of solutions, because the required techniques require substantial domain-specific modifications to existing solvers. Optimal synthesis is the task of producing a solution that not only satisfies the specification but also minimizes a desired cost function.

We present metasketches, a general framework for specifying and solving optimal synthesis problems. Metasketches offer strategic control over the underlying synthesizer by specifying a fragmentation of the search space into an ordered set of classic sketches. We provide two cooperating search algorithms to effectively solve metasketches. A global optimizing search coordinates the activities of local searches, informing them of the costs of potentially-optimal solutions as they explore different regions of the candidate space in parallel. The local searches execute an incremental form of counterexample-guided inductive synthesis to incorporate information sent from the global search.

We present Synapse, an implementation of these algorithms, and show that it effectively solves optimal synthesis problems with a variety of different cost functions. In particular, we show that Synapse can find novel approximations to computational kernels that achieve speed-ups of between 1.6x and 35x without hardware support.

4.4 Stochastic approximations for Stochastic Model Checking

Luca Bortolussi (University of Trieste, IT)

Main reference L. Bortolussi, J. Hillston, "Model checking single agent behaviours by fluid approximation," Information and Computation, Vol. 242, pp. 183–226, 2015.
URL http://dx.doi.org/10.1016/j.ic.2015.03.002

We will briefly review a recent line of work trying to exploit different types of stochastic approximation (fluid approximation, linear noise approximation, moment closures) to model check a Markov population model against specific classes of properties. In the talk, we will focus mostly on individual properties, specified by CSL with rewards or by DTA.

4.5 Counterexample Explanation by Learning Small Strategies in Markov Decision Processes

Tomas Brázdil (Masaryk University – Brno, CZ)

Main reference T. Brázdil, K. Chatterjee, M. Chmelík, A. Fellner, J. Křetínský, "Counterexample Explanation by Learning Small Strategies in Markov Decision Processes," arXiv:1502.02834v1 [cs.LO], 2015.
URL http://arxiv.org/abs/1502.02834v1

While for deterministic systems, a counterexample to a property can simply be an error trace, counterexamples in probabilistic systems are necessarily more complex. For instance, a set of erroneous traces with a sufficient cumulative probability mass can be used. Since these are too large objects to understand and manipulate, compact representations such as subchains have been considered. In the case of probabilistic systems with non-determinism, the situation is even more complex. While a subchain for a given strategy (or scheduler, resolving non-determinism) is a straightforward choice, we take a different approach. Instead, we focus on the strategy – which can be a counterexample to violation of or a witness of satisfaction of a property – itself, and extract the most important decisions it makes, and present its succinct representation. The key tools we employ to achieve this are (1) introducing a concept of importance of a state w.r.t. the strategy, and (2) learning using decision trees. There are three main consequent advantages of our approach. Firstly, it exploits the quantitative information on states, stressing the more important decisions. Secondly, it leads to a greater variability and degree of freedom in representing the strategies. Thirdly, the representation uses a self-explanatory data structure. In summary, our approach produces more succinct and more explainable strategies, as opposed to e.g. binary decision diagrams. Finally, our experimental results show that we can extract several rules describing the strategy even for very large systems that do not fit in memory, and based on the rules explain the erroneous behaviour.

4.6 Approximate Computing on Unreliable Silicon

Andreas Peter Burg (EPFL – Lausanne, CH)

Joint work of Burg, Andreas; Karakonstantis, Georgios; Constantin, Jeremy; Teman, Adam

Approximate computing refers not only to approximating complex computations and algorithms with less complex ones, but can also be the basis for providing robustness against errors due to reliabilities of the underlying hardware. In this talk, we consider two types of hardware failure: timing errors and reliability issues in memories. We describe their impact and critically discuss their potential and issues in the context of approximate computing. We show that tolerating timing errors is particularly tricky, while errors in memories are more straightforward to model and exploit. For the latter, we also point out strategies for testing and quality assurance of unreliable hardware and we mention algorithm techniques to reduce the impact of errors on quality.

4.7 Approximate Overview of Approximate Computing

Luis Ceze (University of Washington – Seattle, US)

Motivation for approximate computing. Overview of approximate computing techniques from language to hardware.

4.8 Programming with Numerical Uncertainties

Eva Darulova (MPI-SWS – Saarbrücken, DE)

Numerical software, common in scientific computing or embedded systems, inevitably uses an approximation of the real arithmetic in which most algorithms are designed. Finite-precision arithmetic, such as fixed-point or floating-point, is a common and efficient choice, but introduces an uncertainty on the computed result that is often very hard to quantify. We need adequate tools to estimate the errors introduced in order to choose suitable approximations which satisfy the accuracy requirements. I will present a new programming model where the scientist writes his or her numerical program in a real-valued specification language with explicit error annotations. It is then the task of our verifying compiler to select a suitable floating-point or fixed-point data type which guarantees the needed accuracy. I will show how a combination of SMT theorem proving, interval and affine arithmetic and function derivatives yields an accurate, sound and automated error estimation which can handle nonlinearity, discontinuities and certain classes of loops.

4.9 The dual value of Probabilistic Abstract interpretation

Alessandra Di Pierro (University of Verona, IT)

Probabilistic Abstract Interpretation is a framework for program analysis that allows us to accommodate probabilistic properties and properties of probabilistic computations. We illustrate the dual value of this framework for both deducing and inferring probabilities. More specifically we show the use of PAI for both static analysis and statistical reasoning. The basic ingredient of the PAI framework that makes this possible is the notion of Moore-Penrose pseudo inverse and its least-square approximation property.

Suppose that we want to analyse a program to check whether it is secure up to a given level of accuracy. We can use probabilistic abstract interpretation as follows:

- Define mathematically what ‘secure’ means (e.g. as a probabilistic relation)
- Consider the semantics of the program restricted to this property (abstraction)
- Construct the Moore-Penrose generalised inverse of the abstraction in order to identify an ideal concrete system that satisfies the property up to the fixed accuracy.

Note that the concrete probabilities defining the concrete ideal system are just assumed and may have no relation with the real world.

Now suppose that we have some observations y at hand and we want to use them in order to define an ideal concrete system which is closer to the real one. To this purpose we can use probabilistic abstract interpretation as a linear statistical model in the way explained below:

- Consider the space V of all possible ideal concrete semantics (abstract domain)
- Define a mapping X from V to all possible observations (design matrix)
- Construct the MP generalised inverse of X in order to obtain the best estimate b of the concrete semantics that realises y.

Note that this is nothing else than the application of the Gauss-Markov theorem for linear regression in its simplest version.

4.10 Termination of Probabilistic Programs

Luis María Ferrer Fioriti (Universität des Saarlandes, DE)

Main reference L. M.Ferrer Fioriti, H. Hermanns, “Probabilistic Termination: Soundness, Completeness, and Compositionality,” in Proc. of the 42nd Annual ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages (POPL’15), pp. 489–501, ACM, 2015.
URL http://dx.doi.org/10.1145/2775051.2677001

The talk is an overview of the ranking supermartingale framework to prove almost sure termination of probabilistic programs.

4.11 Quality-Energy Aware System Design

Andreas Gerstlauer (University of Texas – Austin, US)

URL http://www.ece.utexas.edu/~gerstl/research.html#approximate

Approximate computing has emerged as a novel paradigm for achieving significant energy savings by trading off computational precision and accuracy in inherently error-tolerant applications. This introduces a new notion of quality as design parameter. Such approaches will only be successful, however, if quality can be guaranteed and design spaces can be efficiently explored. While ad-hoc solutions have been explored, systematic approaches are lacking. We have been investigating such quality-energy aware system design. At the hardware level, design strategies for synthesis of approximate arithmetic and logic circuits, including adders and multipliers demonstrate existence of a large design space of Pareto-optimal solutions. Such building blocks in turn form the basis for high-level synthesis of hardware and software into approximate datapaths of custom or programmable processors under a range of statistical quality constraints. Finally, at the system level, we envision a key question to be how to address the problem of quality-energy aware mapping and scheduling of application tasks onto general, quality-configurable system platforms.

4.12 Probabilistic Programming Process Algebra

Jane Hillston (University of Edinburgh, GB)

Joint work of Hillston, Jane; Georgoulas, Anastasis; Sanguinetti, Guido
Main reference A. Georgoulas, J. Hillston, D. Milios, G. Sanguinetti, "Probabilistic Programming Process Algebra," in Proc. of the 11th Int'l Conf. on Quantitative Evaluation of Systems (QEST'14), LNCS, Vol. 8657, pp. 249–264, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-10696-0_21

Formal modelling languages such as process algebras are effective tools in computational biological modelling. However, handling data and uncertainty in these representations in a statistically meaningful way is an open problem, limiting their usefulness in many real biological applications. In contrast, the machine learning community have recently proposed probabilistic programming as a way of expressing probabilistic models in a language which incorporates distributions and observations, and offers automated inference to update the likely distribution over values given the observations.

I will present work which seeks to combine these approaches allowing formal mechanistic models which encompass uncertainty, observations and inference.

4.13 On Quantification of Accuracy Loss in Approximate Computing

Ulya R. Karpuzcu (University of Minnesota – Minneapolis, US)

Emerging applications such as R(ecognition), M(ining), and S(ynthesis) suit themselves well to approximate computing due to their intrinsic noise tolerance. RMS applications process massive, yet noisy and redundant data by probabilistic, often iterative, algorithms. Usually the solution space has many more elements than one, rendering a range of application outputs valid, as opposed to a single golden value. A critical step in translating this intrinsic noise tolerance to energy efficiency is quantification of approximation-induced accuracy loss using application-specific metrics. This article covers pitfalls and fallacies in the development and deployment of accuracy metrics.

4.14 Understanding and Analysing Probabilistic Programs

Joost-Pieter Katoen (RWTH Aachen, DE)

We develop program analysis techniques, based on static program analysis, deductive verification, and model checking, to make probabilistic programming more reliable, i.e., less buggy. Starting from a profound understanding from the intricate semantics of probabilistic programs (including features such as observations, possibly diverging loops, continuous variables, non-determinism, as well as unbounded recursion), we study fundamental problems such as checking program equivalence, loop-invariant synthesis, almost-sure termination, and pre- and postcondition reasoning. The aim is to study the computational hardness of these problems as well as to develop (semi-) algorithms and accompanying tool-support. The ultimate goal is to provide lightweight automated means to the probabilistic programmer so as check elementary program properties.

4.15 Computing Reliably with Molecular Walkers

Marta Kwiatkowska (University of Oxford, GB)

Main reference F. Dannenberg, M. Kwiatkowska, C. Thachuk, A. J. Turberfiled, "DNA walker circuits: computational potential, design, and verification," Natural Computing, 14(2):195–211, 2015; pre-print available from project repository.
URL http://dx.doi.org/10.1007/s11047-014-9426-9
URL http://www.veriware.org/bibitem.php?key=DKTT15

DNA computing is emerging as a versatile technology that promises a vast range of applications, including biosensing, drug delivery and synthetic biology. DNA logic circuits can be achieved in solution using strand displacement reactions, or by decision-making molecular robots-so called 'walkers'-that traverse tracks placed on DNA 'origami' tiles. Similarly to conventional silicon technologies, ensuring fault-free DNA circuit designs is challenging, with the difficulty compounded by the inherent unreliability of the DNA technology and lack of scientific understanding. This lecture will give an overview of computational models that capture DNA walker computation and demonstrate the role of quantitative verification and synthesis in ensuring the reliability of such systems. Future research challenges will also be discussed.

4.16 Approximate counting for SMT

Rupak Majumdar (MPI-SWS – Kaiserslautern, DE)

♯SMT, or model counting for logical theories, is a well-known hard problem that generalizes such tasks as counting the number of satisfying assignments to a Boolean formula and computing the volume of a polytope. In the realm of satisfiability modulo theories (SMT) there is a growing need for model counting solvers, coming from several application domains (quantitative information flow, static analysis of probabilistic programs). We show a reduction from an approximate version of ♯SMT to SMT.

We focus on the theories of integer arithmetic and linear real arithmetic. We propose model counting algorithms that provide approximate solutions with formal bounds on the approximation error. They run in polynomial time and make a polynomial number of queries to the SMT solver for the underlying theory. We show an application of ♯SMT to the value problem for a model of loop-free probabilistic programs with nondeterminism.

4.17 Smoothed Model Checking: A Machine Learning Approach to Probabilistic Model Checking under Uncertainty

Dimitrios Milios (University of Edinburgh, GB)

Probabilistic model checking can provide valuable insights on the properties of stochastic systems. In many application fields however, it is not always possible to accurately identify some of the parameters of the model in question. It is therefore desirable to be able to perform model checking in presence of uncertainty. We show that the satisfaction probability of a temporal logic formula is a smooth function of the model parameters. This smoothness property enables us to construct an analytical approximation of the satisfaction function by using a well-established machine learning framework for approximating smooth functions. Extensive experiments on non-trivial case studies show that the approach is accurate and several orders of magnitude faster than naive parameter exploration with standard statistical model checking methods.

4.18 Accuracy-Aware Compiler Optimizations

Sasa Misailovic (MIT – Cambridge, US)

Joint work of Martin Rinard, Michael Carbin, Hank Hoffmann, Stelios Sidiroglou, Sara Achour, Zichao Qi, and Sasa Misailovic

Many modern applications (such as multimedia processing, machine learning, and big-data analytics) exhibit a natural tradeoff between the accuracy of the results they produce and the application's execution time or energy consumption. These applications allow us to investigate new, more aggressive optimization approaches.

I present a novel approximate optimization framework based on accuracy-aware program transformations. These transformations trade accuracy in return for improved performance, energy efficiency, and/or resilience. The optimization framework includes program analyses that characterize the accuracy of transformed programs and search techniques that navigate the tradeoff space induced by transformations to find approximate programs with profitable tradeoffs. I will present how we can use this accuracy-aware optimization framework to 1) automatically generate approximate programs with significantly improved performance and acceptable accuracy, and 2) automatically generate approximate functions that maximize energy savings when executed on approximate hardware platforms, while ensuring that the generated functions satisfy the developer's accuracy specifications.

4.19 Intuitors, Computers and Validators: Towards Effective Decision-Making Systems

Ravi Nair (IBM TJ Watson Research Center – Yorktown Heights, US)

Traditional computer systems are designed for applications such as transaction processing and physical simulations, largely using systematic algorithms with reliable computation and data movement. Machines are increasingly being asked to produce actionable results to large scale problems for which neither the data nor the available contextual information is 100% reliable. Approximate computing has been making significant headway towards better resource utilization for such new workloads, but the machines executing them still largely maintain the logical and deliberate nature of computer systems designed for traditional workloads. In several respects, today's computers are analogous to the slow, logical, and deliberate System 2 mode of human thought as described in the Nobel Laureate, Daniel Kahneman's book, "Thinking, Fast and Slow." We postulate that Kahneman's System 1 mode of thought, characterized by fast, intuitive, and energy-efficient decision making, suggests a new type of machine for new workloads, which we call an intuitor, which is different from a traditional computer. The incorporation of a validator which monitors the validity of the decision produced by an intuitor, allows the system to tolerate extreme forms of approximation, employing new types of devices and non-traditional architectures, in the design of intuitors. This talk will outline the symbiotic role of intuitors, computers, and validators in future decision-making systems.

4.20 Error Resilient Systems and Approximate Computing: Conjoined Twins Separated at Birth

Karthik Pattabiraman (University of British Columbia – Vancouver, CA)

Main reference A. Thomas, K. Pattabiraman, "Error detector placement for soft computation," in Proc. of the 43rd Annual IEEE/IFIP Int'l Conf. on Dependable Systems and Networks (DSN), pp. 1–12, IEEE, 2013.
URL http://dx.doi.org/10.1109/DSN.2013.6575353

The fields of approximate computing and error resilient systems have evolved independently, though they have a shared origin, namely how to ensure correctness in the presence of hardware faults ? In this talk, I will examine the similarities and differences between the two fields and how we can learn from each other. I will also present an example of a system that my students and I have worked on that attempts to bridge the gap between the two areas. I will conclude by presenting future challenges and opportunities in this area.

4.21 ACCEPT: We Built an Open-Source Approximation Compiler Framework So You Don't Have To

Adrian Sampson (University of Washington – Seattle, US)

Main reference A. Sampson, "Probabilistic Programming," Lecture notes, 2015.
URL http://adriansampson.net/notes/5d3mpq5r6ab0/
Main reference ACCEPT – An Approximate Compiler, Documentation.
URL http://accept.rocks/

Building and evaluating a new technique for approximate computing involves a lot of boring infrastructure work that can be far afield from the core of your work. You need a program annotation system to choose what to approximate, and you will want help writing annotations. You will want to tune each benchmark to take the best advantage of your new technique, and you will need to evaluate the final results on new inputs. If your technique works at a coarse grain, like a hardware accelerator does, you will need to search for large approximate regions to maximize the technique's effectiveness.

If every researcher continues to plod through these same steps independently, the community will waste a tragic amount of time in aggregate. As a fledgling research community, we need to collaborate on common infrastructure to build momentum in the field.

ACCEPT, the Approximate C Compiler for Energy and Performance Trade-offs, is an open-source framework that includes all the boring parts of building and evaluating an approximation technique. It has an annotation system, compiler feedback for the programmer, region inference, an auto-tuner, and Pareto frontier evaluation output. It comes with a suite of C and C++ benchmarks ready to run through the system. The source and documentation for ACCEPT are available now at http://accept.rocks/.

4.22 Approximate Storage

Karin Strauss (Microsoft Corporation – Redmond, US)

Joint work of Sampson, Adrian; Guo, Qing; Nelson, Jacob; Strauss, Karin; Ceze, Luis; Malvar, Henrique
Main reference A. Sampson, J. Nelsen, K. Strauss, L. Ceze, "Approximate Storage," in Proc. of the 46th Annual IEEE/ACM Int'l Symp. on Microarchitecture, pp. 25–36, ACM, 2013; pre-print available from company's repository.
URL http://dx.doi.org/10.1145/2540708.2540712
URL http://research.microsoft.com/apps/pubs/default.aspx?id=207379
Main reference Q. Guo, K. Strauss, L. Ceze, R. Malvar, "High-Density Image Storage Using Approximate Memory Cells," to appear in Proc. of the 21th ACM Int'l Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS'16); pre-print available from company's repository.
URL http://research.microsoft.com/apps/pubs/default.aspx?id=258662

In this talk, I will present the concept of approximate storage. Certain applications have inherent levels of noise and imprecision in them, yet memories still provide very high fidelity storage. However, scaling these memories to higher density is ever more challenging, and relaxing high fidelity requirements for tolerant applications may come to the rescue. I will show how to do this in a disciplined manner and report on the benefits of such approach. I will then describe our experience with storing images in approximate storage. If done naively, the quality degradation can be unacceptable. I will present an algorithm that takes importance of encoded bits on output quality into account during the encoding process to appropriately leverage approximate storage. It requires a small modification to an existing algorithm, yet it reduces quality degradation to practically imperceptible levels.

4.23 DNA Storage

Karin Strauss (Microsoft Corporation – Redmond, US)

Joint work of Bornholt, James; Lopez, Randolph; Carmean, Douglas M.; Ceze, Luis; Seelig, Georg; Strauss, Karin
Main reference J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, K. Strauss, "A DNA-Based Archival Storage System," to appear in Proc. of the 21th ACM Int'l Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS'16); pre-print available from company's repository.
URL http://research.microsoft.com/apps/pubs/default.aspx?id=258661

In this talk, I will describe our project on using a DNA substrate to store digital data. DNA is dense, can be made very durable, and is easy to manipulate. I will explain how data can be stored in DNA, its advantages and challenges, and how to address some of these challenges. In specific, I will provide an overview of how to implement random access by leveraging existing protocols very common in life sciences research, and one way to encode digital data in DNA to improve its reliability while keeping overheads low.

4.24 Quantifying Program Differences

Willem Visser (Stellenbosch University – Matieland, ZA)

We will show to calculate the difference between two programs using Probabilistic Symbolic Execution. More specifically we will show that one can count the number of solutions to a path condition during symbolic execution and use this to calculate the percentage of inputs on which two programs give different outputs. A brief example will be given of how this work to analyse program mutations.

4.25 On a Framework for Quantitative Program Synthesis

Herbert Wiklicky (Imperial College London, GB)

Main reference H. Wiklicky, "Program Synthesis and Linear Operator Semantics," in Proc. of the 3rd Workshop on Synthesis (SYNTH'14), EPTCS, Vol. 157, pp. 17-33, 2014.
URL http://dx.doi.org/10.4204/EPTCS.157.6

Arguably most work on the problem of program synthesis is based on various models based in discrete structures, e.g. related to model checking, game theoretic models, combinatorial optimisation, etc. In this talk we aim in recasting program synthesis as a non-linear, continuous optimisation problem. This allows among other things for a smoother integration of non-functional constraints. Initial experiments demonstrate that, maybe surprisingly, it is possible to avoid algebraic reasoning for algebraic problems and replace it entirely by continuous optimisation constraints.

5 Achievements of this Seminar

Participants attending the seminar represented all four themes of the seminar. The program consisted of (1) tutorials, which introduced each of the main areas to all of the participants on the first day of the seminar, (2) 15-minute individual talks, which presented current research of the participants during the remaining days, (3) breakout sessions, during which the participants had an opportunity to discuss in more details specific points of interest, and (4) a panel, which discussed the main challenges and interactions between the areas.

Relations between the Areas. The participants identified probability and probabilistic reasoning as the underlying basis of all four areas. Figure 1 presents the main interactions between the areas[1]. For instance, some of the existing and anticipated interactions include:

- Probabilistic model checking, with its ability to establish whether a desired property of a probabilistic system holds, can be used to (1) verify the properties of approximate hardware and software systems against the formal specifications of their desired behavior, and (2) verify probabilistic assertions in probabilistic programs. In addition, probabilistic

[1] Figure 1 was compiled by Luis Ceze.

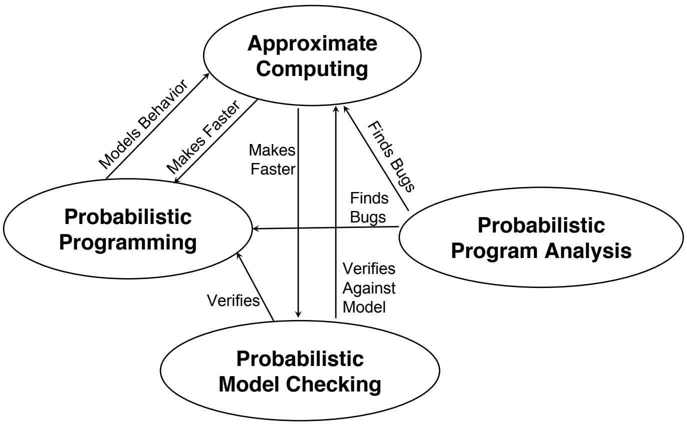

Figure 1 The main identified interactions among the areas.

model checking techniques based on dynamic programming have the flavor of any-time computation, and naturally lend themselves to approximate computation.

- Quantitative program analysis, such as probabilistic symbolic execution, can be used to (1) help find bugs and analyze properties of approximate software, which often implements randomized and/or probabilistic algorithms, and (2) improve testing of probabilistic inference engines and provide alternative strategies for computing results for some classes of probabilistic inference problems.
- Probabilistic programming, with its ability to represent complicated probabilistic models as computer programs and automate inference, can, in principle, represent a basis for specifying rich models of approximate software and hardware systems and enable Bayesian reasoning about the properties and self-adaptability of such systems.
- Approximate computing, with its ability to find efficient architecture and system level approximations for many emerging application domains, including probabilistic inference, has the potential to speed up various common inference tasks in probabilistic computing. But as it requires qualitative assurance of accuracy, it represents a potentially fruitful domain for applying systematic probabilistic reasoning studied in the remaining three areas, and thus creating novel expressive, precise, and scalable program/system analyses.

Open Research Questions. During the individual talks and the breakout sessions the participants identified and discussed many open research challenges and potentially fruitful directions, including the following:

- A key challenge for applying probabilistic reasoning to analyze approximate hardware is precise-enough modeling of the underlying phenomena that lead to approximate and/or unreliable results produced by a device. To that end, future research includes (1) selecting an appropriate levels of abstraction for a variety of hardware models and sources of result inaccuracy and (2) exposing inaccuracy and unreliability via appropriate specifications to the software level of the computing stack are open research problems. The approximate computing community, along with researchers from probabilistic programming, probabilistic model checking, and probabilistic verification, all need to develop a cogent specification of what it means to be approximate.
- Specifying and checking quality of approximate programs can be more systematized. Further work on benchmark suites for approximate computing – including specifications of representative inputs, quality metrics, and acceptable tolerance – can improve design

of future approximation and optimization techniques, and provide researchers from other areas with representative programs for testing their analyses. Furthermore, the development of domain-agnostic, standardized quality measures to push the interoperability of approximate computing applications.

- Understanding quality requirements of approximate subcomputations and code-level specifications, such as the frequency and/or magnitude of the errors of approximate subcomputations, can lead to new numerical analysis approaches that take advantage of system-level approximations, while providing theoretical guarantees for the behavior (for instance convergence) of the full algorithm.
- Some software is inherently resilient. For example, many numerical methods (e.g., iterative methods to learn a linear model) are naturally robust to noise. These algorithms offer a special playground for approximate hardware: if their robustness is sufficient to deal with the weak, non deterministic guarantees of an approximate hardware, the latter can be used for a faster and cheaper execution; otherwise the program can fall back to non-approximate hardware. Identifying for which algorithm this pattern can be fruitfully applied can drive a new generation of numerical libraries and pave the way to the definition of design guidelines for extending the approach to other classes of algorithms.
- *Thinking, Fast and Slow* is a best-selling book by Daniel Kahneman which posits humans use two high level modes of thought: "*system 1*", which is a fast and instinctive judgement and "*system 2*", which is computationally demanding and logical. This insight has been discussed in the context of approximate computing, where a cheap, fast to compute *system 1* approximate solution may be enhanced with a quantified confidence measure; the lack of a sufficient confidence on *system 1* results may trigger the use of a more deliberate, expensive, and proof-based *system 2*, which can provide more accurate results and reasons about whether the model uses by *system 1* is sufficient. This two-level pattern for building approximate systems seems promising for a variety of applications.
- Developing verification and abstraction techniques for probabilistic programs is a critical issue. The specification of probabilistic programs, as well as the meaning of correctness in this quantitative domain, have no generally accepted formalization. The semantics of simplified languages (e.g., constraining the input domain or the language operations) has been successfully abstracted into established stochastic models, such as Markov chains or Bayesian networks, inheriting the corpus of techniques developed in that area. However, the abstraction of more complex language constructs is still an open challenge. Furthermore, the generalization of recent results on probabilistic termination have to be investigated for complex probabilistic programming languages.
- Probabilistic programming and probabilistic program analysis share the development of a core of inference techniques. During the seminar, some inference problem arising from probabilistic programming have been efficiently solved using solution space quantification techniques from quantitative program analysis. However, the expressiveness of probabilistic programming goes beyond the current capabilities of quantitative program analysis, pushing for the study of new and more efficient solution space quantification techniques.
- Quantitative information about a program execution can inform program synthesis and repair approaches. Their usage at compiler level can be the basis of program optimization tailored to specific usage profiles. At the application level, quantitative information may guide the developer in representing the impact different code blocks have on the satisfaction of a program requirements, guiding debugging and prioritizing code refinements.

Case Studies. The seminar participants discussed various applications that can be used as inspiration for new research ideas that span multiple areas, in addition to classical application domains previously discussed in the literature. Two new emerging applications that span the spectrum include *self-driving cars* (investigated by several car manufacturers) and *mobile personal assistant programs* (such as Apple Siri, Google Now, and Microsoft Cortana). Both of these applications are characterized by uncertain data (e.g., coming from sensors) and environment (e.g., physical properties of the hardware), and their operation is routinely affected by human interaction.

However, the approaches for developing these applications have different objectives and different complementary expertise of the designers. Self-driving cars require strict certification, which in most cases includes formal verification of various timing and safety properties of the car components. Probabilistic verification, analysis, and control under uncertainty can, in principle, provide required guarantees that these properties hold. For this example, system-level approximations have the potential to help meet timing deadlines, but they need to be rigorously modeled and controlled.

In contrast, the tasks of personal assistant programs, which extract information and provide recommendations/opinions to the user, are considered best-effort computations. These applications typically perform natural language processing, probabilistic inference, and learning, for which guarantees of desirable program properties are welcome, they are typically not required for an end-to-end result quality. Personal assistant programs running on mobile devices therefore have more freedom to select the type and level of approximation, especially using new configurable approximate hardware components that give promise to significantly increase battery life.

Conclusion. The main objective of this seminar has been to discuss approaches to model and enable programs to seamlessly operate on uncertain data and computations. It has brought together academic and industrial researchers from the areas of probabilistic model checking, quantitative software analysis, probabilistic programming, and approximate computing. The discussion, enriched by the heterogeneity of the participants' perspectives, allowed the identification of several intersections among the interests of the four areas and a variety or research challenges that span across their boundaries. We anticipate that these together will contribute to the definition of the shared agenda among the four research communities.

Participants

- Sara Achour
MIT – Cambridge, US
- David Bindel
Cornell University, US
- Mateus Araújo Borges
Universität Stuttgart, DE
- James Bornholt
University of Washington – Seattle, US
- Luca Bortolussi
University of Trieste, IT
- Tomas Brázdil
Masaryk University – Brno, CZ
- Andreas Peter Burg
EPFL – Lausanne, CH
- Luis Ceze
University of Washington – Seattle, US
- Eva Darulova
MPI-SWS – Saarbrücken, DE
- Alessandra Di Pierro
University of Verona, IT
- Luis María Ferrer Fioriti
Universität des Saarlandes, DE
- Antonio Filieri
Imperial College London, GB
- Jaco Geldenhuys
University of Stellenbosch, ZA
- Andreas Gerstlauer
University of Texas – Austin, US
- Lars Grunske
HU Berlin, DE
- Jane Hillston
University of Edinburgh, GB
- Ulya R. Karpuzcu
University of Minnesota – Minneapolis, US
- Joost-Pieter Katoen
RWTH Aachen, DE
- Marta Kwiatkowska
University of Oxford, GB
- Rupak Majumdar
MPI-SWS – Kaiserslautern, DE
- Dimitrios Milios
University of Edinburgh, GB
- Sasa Misailovic
MIT – Cambridge, US
- Subhasish Mitra
Stanford University, US
- Todd Mytkowicz
Microsoft Corporation – Redmond, US
- Ravi Nair
IBM TJ Watson Res. Center – Yorktown Heights, US
- Karthik Pattabiraman
University of British Columbia – Vancouver, CA
- Adrian Sampson
University of Washington – Seattle, US
- Karin Strauss
Microsoft Corporation – Redmond, US
- Willem Visser
Stellenbosch University – Matieland, ZA
- Herbert Wiklicky
Imperial College London, GB

Report of Dagstuhl Seminar 15492

Computational Metabolomics

Edited by

Sebastian Böcker[1], Juho Rousu[2], and Emma Schymanski[3]

1 **Universität Jena, DE,** sebastian.boecker@uni-jena.de
2 **Aalto University, FI,** juho.rousu@aalto.fi
3 **Eawag – Dübendorf, CH,** emma.schymanski@eawag.ch

Abstract

The Dagstuhl Seminar 15492 on Computational Metabolomics brought together leading experimental (analytical chemistry and biology) and computational (computer science and bioinformatics) experts with the aim to foster the exchange of expertise needed to advance computational metabolomics. The focus was on a dynamic schedule with overview talks followed by breakout sessions, selected by the participants, covering the whole experimental-computational continuum in mass spectrometry, as well as the use of metabolomics data in applications. A general observation was that metabolomics is in the state that genomics was 20 years ago and that while the availability of data is holding back progress, several good initiatives are present. The importance of small molecules to life should be communicated properly to assist initiating a global metabolomics initiative, such as the Human Genome project. Several follow-ups were discussed, including workshops, hackathons, joint paper(s) and a new Dagstuhl Seminar in two years to follow up on this one.

Seminar November 29 to December 4, 2015 – http://www.dagstuhl.de/15492
1998 ACM Subject Classification J.3 Life and medical science
Keywords and phrases algorithms, bioinformatics, cheminformatics, computational mass spectrometry, computational metabolomics, databases
Digital Object Identifier 10.4230/DagRep.5.11.180
Edited in cooperation with Céline Brouard and Franziska Hufsky

1 Executive Summary

Sebastian Böcker
Juho Rousu
Emma Schymanski

Metabolomics has been referred to as the apogee of the omics-sciences, as it is closest to the biological phenotype. Mass spectrometry is the predominant analytical technique for detecting and identifying metabolites and other small molecules in high-throughput experiments. Huge technological advances in mass spectrometers and experimental workflows during the last decades enable novel investigations of biological systems on the metabolite level. But these advances also resulted in a tremendous increase of both amount and complexity of the experimental data, such that the data processing and identification of the detected metabolites form the largest bottlenecks in high throughput analysis. Unlike proteomics, where close co-operations between experimental and computational scientists have been established over the last decade, such cooperation is still in its infancy for metabolomics.

The Dagstuhl Seminar on Computational Metabolomics brought together leading experimental and computational side experts in a dynamically-organized seminar designed to foster the exchange of expertise. Overview talks were followed by breakout sessions on topics covering the whole experimental-computational continuum in mass spectrometry.

2 Table of Contents

3 Major topics

3.1 Data exchange

Pieter Dorrestein (University of California – San Diego, US)

Much discussion over the past decade in metabolomics has been around data sharing. Several metabolomics repositories exist. I asked how many people here have gone to those databases and used a dataset. Only three people raised their hands, yet this it the community that is developing tools for analysis of datasets. There are several purposes for databases:

- to capture and share metabolomics knowledge,
- to share data,
- to make chemical knowledge accessible,
- to associate metadata with the chemical knowledge.

Then one can build the computational infrastructure to retrieve metabolomics knowledge. An argument was made that we should build an analysis infrastructure that organizes and visualizes data while capturing the data metadata and computing in a distributive fashion. Future opportunities are:

- creation of living data, where data is transferred to users,
- connection to genomic information,
- assessing in silica approaches for new spectral matching functions/algorithms with a common set of LC-MS data sets (e.g. 100,000 data sets),
- relaying new information obtained with new tools to users, rather than each user doing their own search.

3.2 Searching in Structure Databases

David Wishart (University of Alberta, Edmonton, CA)

The presentation described the current state of searching for compounds in metabolic databases. There are three kinds of databases: general compound databases, public repositories and spectral databases. A major problem with the general compound databases is that they do not provide species or functional information regarding the compounds. As a result, there are now a growing number of species-specific compound databases.

This presentation also reviewed some of the key challenges facing metabolomics with regard to molecular structures searching. In particular:

- While the size of the spectral databases is growing, the actual number of compounds is not. How to increase these numbers?
- Only a small fraction of currently known metabolites have (or will have) reference LC-MS spectra. This is a real knowledge dichotomy!
- Even if we would product MS spectra for all know compounds, we would likely only identify 30% of the compounds in untargeted LC-MS. What are we missing?

I discussed some possible solutions to these, including:

- the development of compound libraries and compound exchanges,
- the development of MS/MS production tools like CFM-ID or CSI:FingerID,
- the development of structure/metabolite prediction tools.

3.3 Incorporating Experimental Knowledge

P. Lee Ferguson (Duke University, Durham, US)

Experimental knowledge can be used primarily in two ways to identify compounds in non-targeted high resolution mass spectrometry workflows. First, data such as chromatography retention time, ionization performance, and metadata such as reference count and chemical production volume can be used to refine compound identification, after data acquisition. Second, experimental data such as fates or effects of compounds can be used to prioritize data features for subsequent identification. Frontiers such as LCxLC and X-ray crystallography were introduced as future directions.

3.4 Using retention index information of an orthogonal filter for compound identification in GC/MS analysis

Tom Wenseleers (KU Leuven, BE)

In this talk I gave an overview of the potential of using retention index information for compound identification in GC/MS analysis, especially when combined with other pieces of orthogonal information, including electron impact and chemical ionization spectra, *in silico* predicted EI spectra and mass and isotope abundance information.

I provided several examples of compounds where retention index was really critical for correct identification, even if EI mass spectral fragments and mass could be measured with perfect accuracy. I then pointed out the potential of building combinatorial libraries with compounds that are biologically plausible and adding *in silico* predicted EI spectra and retention indices. A proof-of-concept was provided where this method was able to correctly identify ca. 10000 methylalkanes. I finished by discussing database requirements and the need for standardized data formats to include and more retention index information.

3.5 Utilization of retention time in LC-MS

Michael A. Witting (Helmholtz Zentrum, München, DE)

A lot of effort is made to analyze MS, MS2, MS3 ... spectra, but orthogonal information like separation dimension or ion mobility are often neglected. However to improve identification of unknown or verification of known molecules they have to be incorporated. To facilitate data sharing a novel retention time indexing for RP-LC-MS was presented. This indexing system will potentially allows integrated analysis of RTI data from different sources compiled on similar systems. Additionally, *de novo* prediction of retention times using different published methods was discussed. Several limitations have been identified, which have to be tackled by the community. Lastly, ion mobility as orthogonal method was presented.

4 Generating Spectra in silico

4.1 In Silico Mass Spectral Identification

Tobias Kind (University of California – Davis, US)

In silico methods for mass spectrometry can be used to calculate spectra directly from chemical structures. Traditionally spectra had to be acquired by experimental measurements only, now purely computational methods can be used. This includes *ab initio* methods, machine learning methods, reaction based tools and heuristic methods. Their outputs have to be validated and prediction accuracy has to be tuned for better performance. In the future it will be possible to generate millions of mass spectra (hopefully highly accurate), which then will lead to the following problem: the curse of similarity and potential database poisoning with millions of similar spectra.

4.2 Competitive Fragmentation Modeling

Felicity Allen (University of Alberta, Edmonton, CA)

Existing methods for spectrum prediction generally produce far more peaks than actually occur in a measured spectrum. Competitive Fragmentation Modeling (CFM) is a method that we propose to predict fewer peaks that are more likely to occur. It uses a probabilistic, generative model of the fragmentation process. Parameters of the model are learned from data using expectation maximization. The method has recently been extended for use with EI-MS. Empirical results show that the method outperforms existing computational tools, but is still inferior to actually measuring the spectrum. Despite this short-coming, actual measurements are often costly or infeasible, and so this methods offers an important alternative.

5 Breakout Groups

5.1 Spectral Simulation

The discussions on spectral simulation started with a survey of who uses what: CFM-ID, QC (quantum chemical)-EI-MS, CSI:FingerID, Mass Frontier, ACD MS Fragmenter, HAMMER, manual interpretation, or a combination of all were mentioned. It was established that mass spectral simulation software needs to accurately predict fragment ions and their peak abundances. Most software produce different fragments and although better ranking results are achieved with e.g. CFM–ID, the fragments are not always “chemically sensible” and in this sense Mass Frontier is often more accurate because it makes use of reaction chemistry from the reference literature. The quantum chemical simulation of Grimme (QC-EI-MS) is promising and theoretically extendable to ESI but because of the complexity of the computational tasks, the quantum chemical community needs to be engaged to solve this. It was discussed whether CFM–ID could “learn” rearrangements, but it needs the knowledge in advance to do this; these cannot be exported from Mass Frontier. Toolkits used included RDkit (C++/python) ChemAxon (free academic), CDK (limited reaction capabilities) – having an active development community behind is essential. The need for more experimental data was discussed, because more data could be used to improve modelling accuracy, once large enough validation sets are available. It was debated whether the Markov approach behind CFM–ID could be used to train intensities for some of the other *in silico* fragmenters. Last ideas included treating the mass spectrum as a picture (picture recognition algorithm) and whether mass spectral data should be uploaded to http://www.kaggle.com (a platform for data prediction competitions) to get very good machine learners working on mass spectra.

5.2 Next Generation Computational Methods

The breakout group on next generation computational methods covered several topics. A debate about identification measures covered whether the current scores for *in silico* fragmenters are sufficient in separating the true from false matches and whether the score should aim to pick the best candidate or rather show how good the prediction is, also considering top K instead of top 1 (see also “Statistics”, below). The “Percolator approach” was also discussed.

The next topic covered joint identification, using the presence of other substances to elevate the ranks of “unknowns” with prior evidence, using mass differences and also clustering by using multiple measurements as training sets to perform machine learning. Estimates included requiring half the number of samples for the number of metabolites under investigation (i.e. under 1000 samples for typical cases).

Finally, discussions ended with substances that are not in the databases and using predicted transformations to help find potential candidates via biotic and abiotic reactions. The presence of peptides, oligonucleotides, sugars and homologue series were also discussed, including the potential to run all small poly-peptides, potentially up to 8, and add them to the Global Natural Product Social Networking (GNPS) library. Discussions ended on a summary figure from GNPS that showed that there is a lot of “dark matter” remaining and very few known annotations, many of the unknowns are singletons.

5.3 Metadata and common input/output formats

The breakout group on metadata focused on what types of metadata would need to be reported for a given study for it to be useful and discussed resurrecting an old SepML standard using controlled vocabulary from existing ontologies. A large number of action points were made, especially involving vendors and Proteowizard, to enable export of given parameters into the open format. Points to discuss in the future remained most recent separation advances: 2D LC and GC (liquid and gas chromatography) as well as ion mobility.

The group on common input/output formats discussed the need to explore common parameters and formats between most software for small molecule identification. Two different use cases evolved: development (simple text-based format, e.g. MGF, Mascot Generic Format) versus pipeline integration once developed (fancy mzML-type format for machine-readable properties). Software-specific parameters can remain flexible. The ability of mzML to support structures may be a limitation with this format. Outputs in CSV files with common column headers or SDFs with common tags were discussed; developers should not rely on a certain order in the CSV for maximum flexibility. Some discussions on potential test data were made. These discussions will continue beyond Dagstuhl.

5.4 Integrative Omics

The breakout group on "integrative omics" discussed that the correlations between the different omics levels are complex and the integration of metabolomics is poor, with no computationally-feasible way to connect the layers. Several issues were discussed to address the lack of interaction information between metabolites and genes/proteins, such as enzyme reaction models, systematic studies of metabolite-protein binding (technically difficult to find), collation of existing knowledge in a protein–metabolite–interaction database in a machine-readable way, as well as computational methods needed to find novel pathways and interactions between different levels (text-mining?).

The combination of transcriptomics with metabolomics was discussed, rather than pure mapping, as this is more orthogonal that proteomics/metabolomics. This could be used to find the most interesting sites in the networks and possibly even help build the network if one could differentiate the data sufficiently. However, this may be hindered by different time-scales as the metabolome changes extremely fast. Finally, correlation feature-based instead of identification-based approaches were mentioned.

5.5 The Dark Matter of Metabolomics

The breakout group on the dark matter of metabolomics and in-source fragmentation phenomena had a pretty wide ranging discussion focusing on the relatively low rates of annotation of compounds/features from LC-MS studies using either MS level data, MS/MS data, or infusion data. The consensus was that 30% seems to be an approximate maximum success rate across labs. The need for a gold-standard ground truth dataset was stressed, to evaluate the various steps in the data processing and annotation processes, from peak picking/feature grouping through the final annotation and evaluation. The need for the full utilization of all existing MS data, and supplementing with non-MS data (biology, computation, NMR, etc) was reiterated to try and address the identification of real and reproducible signals.

5.6 Statistics

The statistics breakout group discussed issues that arise when searching in larger (spectral or molecular structure) databases. Currently, only relatively few compounds are identified in an LC-MS run; when more compounds are putatively identified, this will come at the price of more bogus identifications. This is independent of the fact whether we are searching in a large spectral library, or a large molecular structure database. To this end, scores have to be introduced that express a methods "confidence" that a certain identification is correct. Beyond that, False Discovery Rates (q-values, p-values) would be very helpful to navigate the putative identifications and to find reasonable thresholds of what to accept and what to reject, similar to Shotgun Proteomics. We also discussed the problem of p-value corrections for repeated testing. Finally, we discussed how to combine orthogonal information for compound identification into a single, statistically meaningful measure.

5.7 Metabolite Prediction

The metabolite prediction breakout group discussed two approaches to metabolite prediction:

1. iteratively: start from a set of known compounds, predict, confirm the existence and use this information to refine predictions
2. databases: generate predicted metabolites from large sets of known compounds and filter these "on the fly" – with the risk of combinatorial explosion

The consensus was that a combination of both approaches would be the most practical. As only a fraction of the metabolites in a metabolic network are observed, multiple prediction steps are need to be applied before a path can be confirmed, adding to the combinatorial explosion issue. On the other hand instruments are becoming more sensitive and larger fractions of (predicted) metabolites can be expected to be seen.

Big differences exist in the amount of data available in different "domains of metabolism". In some domains there is enough data to train probabilities (drugs), while in other domains data is scarce and rules are more literature based. In the case of gut transformations rules may represent what goes into a microbe and what comes out, rather than substrates and products of an enzyme. The same may be true for environmental applications.

In addition to empirical or trained likelihoods of biotransformation, kinetic parameters (from simulations) and thermodynamic parameters (which can be calculated) are useful additional parameters to evaluate and prune predicted networks.

5.8 Data visualization

The data visualization group discussed the visualization of complex data in a biological context. Interactive visualization allowing the navigation and exploration of data, going back and forth between the data and the outcomes, was a main topic. The output devices were to be "papers"/software/web apps. Another visualization challenge is looking at the large "lists" of metabolite structures, for instance the hierarchical clustering of metabolite structures in MetFragBeta, also shown in Figure 2 of Schymanski *et al.* 2014. Molecules in chemical space can also be plotted in a PCA format using chemical descriptors, as done in Figure 4 from Kuhn *et al.* 2009.

5.9 The CASMI contest

Steffen Neumann (Leibniz Institute of Plant Biochemistry – Halle, DE)

This breakout session discussed the Critical Assessment of Small Molecule Identification (CASMI) contest, founded in 2012 (http://www.casmi-contest.org). The protein equivalent, CASP, has many more participants but took several years to establish and receives considerable funding each year to run the contest. Several suggestions for future CASMIs were discussed. Participants requested raw data in addition to peak lists, with future peak lists to be provided as MGF as a new standard format for identification tools, with challenges submitted to MassBank. A “spectrum-only” category was discussed, where common candidate lists could be provided and no additional scoring criteria would be allowed, to focus on only *in silico* fragmentation techniques. A detailed description of the analytical conditions (chromatography, mass spectrometry) should be provided. The participants also indicated that they would like a CASMI workshop to discuss the results after closure of the contest; the current “outlet” is in the form of publications, with mixed success. A workshop is under consideration for the 2016 contest. Ideas for future CASMIs included a staged contest (automatic approaches first, results are then published on the website and then manual users have a few more weeks), assigning manual users a sub-category of automatic categories, to enable bigger automatic datasets for statistical robustness, and a “whole box” category where all information sources are allowed. Nuclear Magnetic Resonance spectroscopy was discussed as a new category, as there have been interesting developments recently. The idea of a GNPS/CASMI continuous evaluation dataset was also received positively and there are several challenges (unsolved) available on GNPS already.

5.10 Workflows

The workflow breakout group discussed standardized formats (see also Section 5.3) and that mzTab and mzML would be the potential file types to incorporate all information needed. Participants were strongly encouraged to pass on their ideas for standardization to the Proteomics Standards Initiative (PSI) and ask them to integrate them (and also participate in the initiative). The Spring PSI meeting (April 2016, Ghent) would be an opportunity for this. There were some additional discussions on the contents of the standards as well. Finally, although many pipelines try to get an “all in one” workflow, it was discussed about whether to split workflows into parts, with the large divide (everything before you start to work with statistics) and (after).

5.11 Feature Finding, Quantification, Labelling

Several topics merged into one breakout session. The computational challenges of quantification were discussed, including

- finding all features is challenging (needs to be more flexible/robust, e.g. slow-release substances, presence of m/z and intensity shifts, physical interferences).
- summing the signal to quantify.
- feature alignment across samples is considered essentially solved.
- still no clear idea what is the best normalization method, as this is dependent on experimental design.
- that experimental data contains no real ground truth, but while synthetic data is not appreciated by experimentalists, this is essential for computational people.
- reference datasets are available on the CompMS website.

From the experimentalists point of view, concentrations/quantification is needed to translate detected metabolites to the biology; quantification can be used to model metabolic networks and see fluxes. Instrument ionization is complex and formation of ions varies greatly with structure. Internal standards (preferably isotopically-labelled) are needed; at least one per compound class. Standard additions also possible. The solvent composition can have a huge influence on signal intensities, while the influence of acidity and polarity was also discussed. Questions included whether to sum intensities from all adducts, or remove/ignore smaller signals, how to extract response factors from runs and using ion current measurements to correct for ESI spray fluctuation. Can adduct species be predicted? Labelling experiments can yield even more information, including qualitative and quantitative flux measurements and thus tracking origin and fate of metabolites, yet over 65 % of signals remain unidentified despite labelling proving they have biological origin – see Section 5.5.

6 Hands-on Sessions

A number of small hands-on sessions were run during the meeting. The environmental and xenobiotic session on Tuesday discussed data from different sources in detail and the surprising complementarity observed in the production volume and patent data. At the same time, a breakout on the SPectraL hASH (SPLASH) introduced this concept and determined that these are now google-searchable. One participant now has a roadmap to contribute his substances to MassBank, using MetShot and RMassBank. On the last day, a software demonstration and feedback session was run across the whole morning and was enjoyed by all participants with very honest and constructive feedback and discussions about different approaches.

7 Wrap-ups

The seminar wrap-up started with expressions of interest for a commentary/perspectives paper as a partial summary of discussions – over half of the participants were interested and Pieter Dorrestein will take the lead. Focus on metabolomics and the extension to the exposome and small molecule characterization (chemical genomics? chenomics?). Michael Witting advertised a special issue about unknown identification coming up in *J. Chrom. B* (deadline mid 2016). Lee Ferguson announced the Nontarget 2016 conference in Switzerland, May 29 to June 3. A couple of new ideas such as a society for small molecule characterization or a new open source journal were considered unlikely to get off the ground, but alternative meetings such as in conjunction with the Metabolomics Society conference were considered positively. All participants indicated that they had enjoyed the meeting and would come again; none raised their hand for the opposite. The seminar wrap-up concluded with two main questions:

1. Where do we want to be in a year?
   Establishment of benchmark datasets and standard in/out data structure, improved data and spectral sharing as well as using bioboxes for modular workflows.
2. How to we encourage more people?
   Offer machine learning challenges, expose students to metabolomics, increase the data availability, improve the community building efforts (with workshops such as this Dagstuhl Seminar) and initiatives such as Computational Mass Spectrometry (CompMS), which has coursework on computational metabolomics and proteomics.

Excursion

The excursion on Wednesday afternoon was to Trier, including a city tour and the Christmas market, before dinner near the cathedral. A good time was had by all.

8 Conclusion

The first Dagstuhl Seminar on Computational Metabolomics was a huge success with positive feedback from all participants. A general observation was that metabolomics is in the state that genomics was 20 years ago and that while the availability of data is holding back progress, several good initiatives are present. The importance of small molecules to life should be communicated properly to assist initiating a global metabolomics initiative, such as the Human Genome project. Several follow-ups were discussed, including workshops, hackathons, joint paper(s) and a new Dagstuhl seminar in two years similar to this one.

The organizers wish to acknowledge the contributions of Tobias Kind, who attended on behalf of Oliver Fiehn, Franziska Hufsky and Céline Brouard who collected and typed the hand-written abstracts as well as all participants for their contributions.

Participants

- Felicity Allen
University of Alberta – Edmonton, CA
- Nuno Bandeira
University of California – San Diego, US
- Sebastian Böcker
Universität Jena, DE
- Corey Broeckling
Colorado State University – Fort Collins, US
- Céline Brouard
Aalto University – Espoo, FI
- Jacques Corbeil
University Laval – Québec, CA
- Pieter Dorrestein
University of California – San Diego, US
- Kai Dührkop
Universität Jena, DE
- P. Lee Ferguson
Duke University – Durham, US
- Franziska Hufsky
Universität Jena, DE
- Gabi Kastenmüller
Helmholtz Zentrum – München, DE
- Tobias Kind
Univ. of California – Davis, US
- Oliver Kohlbacher
Universität Tübingen, DE
- Daniel Krug
Helmholtz-Institut, DE
- Kris Morreel
Ghent University, BE
- Steffen Neumann
IPB – Halle, DE
- Tomas Pluskal
Whitehead Institute – Cambridge, US
- Lars Ridder
Netherlands eScience Center – Amsterdam, NL
- Simon Rogers
University of Glasgow, GB
- Juho Rousu
Aalto University – Espoo, FI
- Emma Schymanski
Eawag – Dübendorf, CH
- Huibin Shen
Aalto University – Espoo, FI
- Christoph Steinbeck
European Bioinformatics Institute – Cambridge, GB
- Michael Stravs
Eawag – Dübendorf, CH
- Ales Svatos
MPI für chemische Ökologie – Jena, DE
- Tom Wenseleers
KU Leuven, BE
- Rohan Williams
National Univ. of Singapore, SG
- David Wishart
University of Alberta – Edmonton, CA
- Michael Anton Witting
Helmholtz Zentrum – München, DE
- Gert Wohlgemuth
University of California – Davis, US
- Nicola Zamboni
ETH Zürich, CH